

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 10.3976 Seconds  
(without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTVTIQNGKEMSTTVSE.....ATVLDKNISSKSTNNPNK 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_76.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	748	100.0	2140	2 F95074	serine proteinase,
2	745	99.6	2144	2 A97942	metalloproteinase,
3	110	14.7	558	2 T18467	hypothetical prote
4	108.5	14.5	1038	2 JCS497	claustrin - chicke
5	104	13.9	665	2 B71609	hypothetical prote
6	103.5	13.8	325	2 T18283	hypothetical prote
7	100	13.4	312	2 G81339	probable membrane
8	99	13.2	211	2 T25911	hypothetical prote
9	97.5	13.0	1345	2 S46817	hypothetical prote
10	97	13.0	988	2 T14188	hypothetical prote
11	95.5	12.8	644	2 T47835	hypothetical prote
12	95	12.7	348	2 I37271	cylicin ii - human
13	95	12.7	535	2 T37189	hypothetical prote
14	95	12.7	622	2 A90570	lipoprotein (impor
15	95	12.7	1397	2 T10466	DNA topoisomerase
16	94.5	12.6	3724	2 T18427	hypothetical prote
17	94	12.6	210	2 T28771	hypothetical prote
18	92.5	12.4	219	2 B72291	hypothetical prote
19	92.5	12.4	540	2 D86432	hypothetical prote
20	91	12.2	253	2 T32879	hypothetical prote
21	91	12.2	385	2 T20410	hypothetical prote
22	90.5	12.1	614	2 A84152	hypothetical prote
23	90.5	12.1	670	2 T28391	ORF MSV230 hypothe
24	90.5	12.1	1332	2 S41552	probable transcrip
25	90.5	12.1	2401	2 T28676	rhoetry protein -
26	89.5	12.0	508	2 B81594	hypothetical prote
27	89.5	12.0	508	2 C72074	hypothetical prote
28	89.5	12.0	508	2 E86549	hypothetical prote
29	89.5	12.0	700	2 S67610	probable membrane

30	89.5	12.0	1016	2 T19006	ankyrin related pr
31	89	11.9	528	2 E96795	unknown protein F2
32	89	11.9	645	2 E89883	conserved hypothet
33	89	11.9	1888	2 T39009	hypothetical prote
34	88.5	11.8	301	2 T33068	hypothetical prote
35	88.5	11.8	371	2 A71683	hypothetical prote
36	88.5	11.8	762	2 G88436	protein T04A8.13 [
37	88.5	11.8	791	2 T24435	hypothetical prote
38	88	11.8	276	2 T23451	hypothetical prote
39	88	11.8	1280	2 G96796	IgA-specific metal
40	87.5	11.7	1702	2 A41859	hypothetical prote
41	87.5	11.7	2523	2 T18477	hypothetical prote
42	87	11.6	635	2 F71621	microtubule-associ
43	87	11.6	2464	1 QRMSP1	probable chloroqui
44	87	11.6	2708	2 T09079	hypothetical prote
45	86.5	11.6	792	2 T49989	hypothetical prote

ALIGNMENTS

RESULT 1

F95074

serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: F95074

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: F95074

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2140 <CUR>

A:Cross-references: GB:AE005672; PIDN:AAK74791.1; PID:g14972117; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0641

Query Match 100.0%; Score 748; DB 2; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 6.3e-49;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRVTVTIQNGKEMSTTVSEDFILPVYKGELEKGYQPDGWEISGFEGKDGAGYVNLK 60  
|||||  
Db 1963 HRVTVTIQNGKEMSTTVSEDFILPVYKGELEKGYQPDGWEISGFEGKDGAGYVNLK 2022  
|||||

Qy 61 DTFIKPVFKTEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 120  
|||||  
Db 2023 DTFIKPVFKTEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 2082  
|||||

Qy 121 KDVTATVLDKNISSKSTNNPNK 144  
|||||  
Db 2083 KDVTATVLDKNISSKSTNNPNK 2106  
|||||

RESULT 2

A97942

metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C:Accession: A97942

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bargett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A97942  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99365.1; PID:g15458138; GSPDB:GNO0174  
C:Genetics:  
A:Gene: prta  
C:Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 745; DB 2; Length 2144;  
Best Local Similarity 99.3%; Pred. No. 1.1e-48;  
Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRVTVTQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSK 60  
DB 1967 HRVTVTQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSK 2026

QY 61 DTFIKPVFKKIEEKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSOKSDST 120  
DB 2027 DTFIKPVFKKIEEKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSOKSDST 2086

QY 121 KDVATVLDKNNISSKSTTNNPNK 144  
DB 2087 KDVATVLDKNNISSKSTTNNPNK 2110

RESULT 3  
T18467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2002  
C:Accession: T18467  
R:Lawson, D.; Bowman, S.; Barrell, B.  
A:Reference number: Z18937  
A:Accession: T18467  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-558 <LAW>  
A:Cross-references: EMBL:AL008970; NID:e1407852; PIDN:CAA15610.1  
C:Genetics:  
A:Map position: 3  
A:Introns: 84/1; 160/1  
A:Note: C0465c

Query Match 14.7%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 0.5;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 17 IVSEEDFILPVY-----KGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPV 67  
DB 60 ILGFEDDILYCYCISQLKQSKER--KADGEEDKYLNAKKLKNLTGFIGNKKSDIFIEL 117

QY 68 FKKI--EEKKEE-----ENKPTFDVSK-KKDNQVNHSQLNE-----SHRK 105  
DB 118 LELLINEEKKEBHIAITLNEKN-TNDIKVKVKNENINENYNNENKDNKDKHVSQHN 176

QY 106 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISSKSTTN 140  
DB 177 EHNINNVNLKKEKYTDIQDKKKHRSLSQKSDSYKKEPFNKRKTSIER-SLSNKRGYDE 235

QY 141 NPNK 144  
DB 236 KTNK 239

RESULT 4  
JC5497  
N:Alaternate names: keratan sulfate proteoglycan  
C:Species: Gallus gallus (chicken)  
C:Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 21-Jul-2000  
C:Accession: JC5497; PC4334; S37561

R:Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A:Title: Claustrian, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JC5497; MUID:94157526; PMID:7906711  
A:Accession: JC5497  
A:Molecule type: mRNA  
A:Residues: 1-1038 <BUR1>  
A:Cross-references: EMBL:X67778; NID:g406318; PIDN:CAA47988.1; PID:g406319  
A:Accession: PC4334  
A:Molecule type: protein  
A:Residues: 79-83; 299-412; 485-502 <BUR2>  
A:Experimental source: brain  
C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nerve  
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F:267-270/Region: cell attachment (R-G-D) motif  
F:112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 14.5%; Score 108.5; DB 2; Length 1038;  
Best Local Similarity 28.6%; Pred. No. 1.3;  
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

QY 11 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFK- 69  
DB 595 KPETKTIVAERDV-----TTKEQLGKSETSEKQASEKQDVKPKVTKESVKVEKKA 646

QY 70 KIEEKEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE-----EHSQKSDTKDV 123  
DB 647 KPEEKKDEKPKKEVSKKEEKLPI---KBEKPKKEDIKKEVKKEKVEKKEAKKEV 702

RESULT 5  
B71609  
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: B71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: B71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-665 <GAR>  
A:Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AACT1925.1; PID:g384524  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0680w

Query Match 13.9%; Score 104; DB 2; Length 665;  
Best Local Similarity 25.9%; Pred. No. 1.8;  
Matches 42; Conservative 29; Mismatches 35; Indels 56; Gaps 10;

QY 1 HRVT-VTTQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEI--SGFEKGDAGYVIN 57  
DB 138 HRQNELNLQSGK-----NEQDI-----NKNKRGKQ-----DISNSNAENKKD----- 174

QY 58 LSKDTFIKPVFKKIEEKE-----EENKPTFD-----VSKKDNQVNHSQLNESH 103  
DB 175 -----VKEGVKELEKKEEKKEKISDDHKVEENKSDDHKVEENKSDDHKVEENKSDDH 228

QY 104 RKEDLQR-EHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 144  
DB 229 KIEEVKKVEEHEDEE-----DKKEKKSNKNKDNENK 261

RESULT 6  
T18283  
hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999



C;Species: Homo sapiens (man)  
cylicin 11 - human

Species: Mycoplasma pulmonis

lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CT1P)  
 C:Species: Mycoplasma pulmonis  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change  
 C:Accession: A90570



A:Status: preliminary  
A:Status: type: DNA  
A:Molecule type: DNA  
A:Residues: 1-622 <KUR>  
A:Cross-references: PID:G14089879; PIDN:CAC13638.1; GSPDB:GNC00153  
A:Experimental source: strain UAB CTIP  
C:Genetics:  
A:Gene: MYPU 4650  
A:Genetic code: SGC3

Query Match	12.7%	Score 95	DB 2	Length 622
Best Local Similarity	25.4%	Pred. No. 7.9		
Matches	34	Conservative 26	Mismatches 50	Indels 24
Gaps	5			

  

QY	18	VSEDFILPVYKGLKGYGFDGWEISGFEGKDGAGVYNLSKDTFFKVPFKTIEEKKE	77
DB	85	LSEDNIIIFSLNINNAKFLQDEF-----VSKDEKPKIKQEIFNSQTE	128
QY	78	ENKPTFDVSKKCD----NPQVHNSQLNES--HRKEDLQREHSQKSDSTKDVTATVLDKN	131
DB	129	Q-KITDNISSEDEKKNPKDNNENNNSSDPQKNDEIQKNNSDKLNDNVQDEKANKENS	187
QY	132	-NISSKSTTNPNK	144
DB	188	SNDSKEKNENTNKK	201

RESULT 15  
Tl0466  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000  
C:Accession: Tl0466  
R:Cheeseman, S.J.  
submitted to the EMBL Data Library, September 1995  
A:Reference number: Zl7031  
A:Accession: Tl0466  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1397 <CHE>  
A:Cross-references: EMBL:X79345; NID:g994807  
C:Genetics:  
A:Gene: topoII  
A:Map position: 14  
C:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hydrolyzing)  
C:Keywords: ATP; DNA binding; isomerase; nucleus

Query Match	12.7%	Score 95	DB 2	Length 1397
Best Local Similarity	23.8%	Pred. No. 20		
Matches	36	Conservative	53	Indels 26
				Gaps 6
Qy	9	NGKEMSTIYSEEDFIL--PYVKGELKRGYQDCGWEISGPRGKDGAGVYVNLKSDTFIKP	66	
Db	1132	NEHIIAGITVKDYLLSMFISLTLEK---VEDLLTQLEKERELEILRNITVETWMLK	1188	
Qy	67	VFKKIEE-----KKEEENKPTFDVSKKDNFPQVNHQSQNLNESHKREDLQREHSQ	115	
Db	1189	DIEKVEEAIEFQRNVELSNREESNK--FKVARKQ-----GPSSMKKKKKKLSDEESE	1241	
Qy	116	---KSOSTKQVATVLDKNNISSKSTNNPN	143	
Db	1242	GGTSDSSSEFLVNTLNKKQTNKKTITSSNN	1272	

RESULT 16  
T18427  
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text\_change 09-Jun-2000

C;Accession: T18427  
R;Lawson, D.; Bowman, S.; Barrel, B.  
submitted to the EMBL Data Library, August 1997  
A;Reference number: Z18935  
A;Accession: T18427  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-3724 <Law>  
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1  
C;Genetics:  
A;Introns: 307/1; 1545/2  
A;Note: C0335C

[illegible]

RESULT 17

T28771  
hypothetical protein E03H12.5 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T28771  
R;Nelson, J.; Wohldmann, P.; Sansone, J.  
submitted to the EMBL Data Library, June 1997  
A;description: The sequence of *C. elegans* cosmid E03H12.  
A;Accession number: Z20520  
A;Accession: T28771  
A;Status: preliminary; translated from GB/EMBL/DDJ  
A;Molecule type: DNA  
A;Residues: 1-210 <NBL>  
A;Cross-references: EMBL:AF002099; PIDN:AA047980.1; GSPDB:GNC00022; CESP:E03H12.  
A;Experimental source: strain Bristol N2, clone E03H12  
C;Genetics:  
A;Gene: CESP:E03H12.5  
A;Map position: 4  
A;Introns: 30/2; 201/3

[illegible]

RESULT 18

B72291  
hypothetical protein - Thermotoga maritima (strain MGB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: B72291  
R:Nelson, K.E.; Clayton, R.A.; Gall, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richard  
C.M.





```
QY 65 K---PVFKKIEKKEENKPTFD-----VSKKKO-----NQVNHSQLNESH 103
Db 595 KENGKVSSEKSKTKVDEAPTNDKLTLSVPEGEKEKDTASSTVTVHNVNKNEIKENG 654
QY 104 RKEDLQR-FEHQKSDSTKDVATVLD-KNNISSKSTTN 140
Db 655 KNEEQDMVEESKTEDSSKDAADAAKOTDGLQDKTAEN 693

RESULT 25
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28676; A45521
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507; MUID:97077455; PMID:8920022
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: EMBL:U36927; NID:gi1041784; PID:gi1041785; PIDN:AAB41263.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KE>
A:Cross-references: GB:M34281

Query Match 12.1%; Score 90.5; DB 2; Length 2401;
Best Local Similarity 27.4%; Pred. No. 79;
Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;

QY 6 TIONGKEMSTI-----VSEDFILPVYKGEKGY---OPDGWEISGFEGKDA 52
Db 924 TKNHEKISDRKNSLKIIOFSESYINDI-KKELEKNVLESQNNNTDINQYLSKIEN 982
QY 53 GYVINLSKDTFKPVFKKIEKKEENKPTFDVSKKKONPQVNHSQLNESHKRLQREE 112
Db 983 IV--NLKLNKIKKIIDKVKYETDETEK-----NNKKINAELSNSEKIITQLKENSLLKE 1035
QY 113 HSQKSDSTKD-----VTATVLDKNNISS 135
Db 1036 COSKIKSTIDDNVSECIKNITNLKTYIVNEKNINNT 1072

RESULT 26
B81594
hypothetical protein CP0281 [imported] - Chlamydomphila pneumoniae (strain AR39)
C:Species: Chlamydomphila pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: B81594
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: B81594
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <REA>
A:Cross-references: GB:AE002161; NID:g7189205; PIDN:AAF38139.1; PID:g718920
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0281

Query Match 12.0%; Score 89.5; DB 2; Length 508;
```

```
Best Local Similarity 24.5%; Pred. No. 16;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 64 IKPVFKKIEKKEENKPTFD-----VSKKKONPQVNHSQLNESHK 105
Db 92 VKGVFKKTPQARPEVSSPRLPSHVQHGORLPGLGFRDRIQKRSENPEADLGKMKRSYSD 151
QY 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139
Db 152 GDLDRVGHDSNEDSTEDSRS---EGGEPSSKSS 182

RESULT 27
C72074
hypothetical protein - Chlamydomphila pneumoniae (strain CW1029)
C:Species: Chlamydomphila pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: C72074
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: C72074
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <ARN>
A:Cross-references: GB:AE001632; GB:AE001363; NID:g4376755; PIDN:AAD18613.1; PID:g437675
A:Experimental source: strain CW1029
C:Genetics:
A:Gene: CPn0473

Query Match 12.0%; Score 89.5; DB 2; Length 508;
Best Local Similarity 24.5%; Pred. No. 16;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 64 IKPVFKKIEKKEENKPTFD-----VSKKKONPQVNHSQLNESHK 105
Db 92 VKGVFKKTPQARPEVSSPRLPSHVQHGORLPGLGFRDRIQKRSENPEADLGKMKRSYSD 151
QY 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139
Db 152 GDLDRVGHDSNEDSTEDSRS---EGGEPSSKSS 182

RESULT 28
E86549
hypothetical protein CPJ0473 [imported] - Chlamydomphila pneumoniae (strain J138)
C:Species: Chlamydomphila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: E86549
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: E86549
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <STO>
A:Cross-references: GB:BA000008; NID:g8978843; PIDN:BAA98679.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPJ0473

Query Match 12.0%; Score 89.5; DB 2; Length 508;
Best Local Similarity 24.5%; Pred. No. 16;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 64 IKPVFKKIEKKEENKPTFD-----VSKKKONPQVNHSQLNESHK 105
Db 92 VKGVFKKTPQARPEVSSPRLPSHVQHGORLPGLGFRDRIQKRSENPEADLGKMKRSYSD 151
QY 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139
```

Db 152 GDLDRVGHSDNEDSTEDSRS---EGGEPSSKSS 182

RESULT 29

S67610

probable membrane protein YDL074c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein D2483

C:Species: *Saccharomyces cerevisiae*

C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002

C:Accession: S67610

R:Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67608

A:Accession: S67610

A:Molecule type: DNA

A:Residues: 1-700 <WAM>

A:Cross-references: EMBL:Z74122; NID:gl431087; PID:e253213; PID:gl431088; GSPDB:GN00004;

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:BRE1; MIPS:YDL074c

A:Cross-references: SGD:S0002232

A:Map position: 4L

C:Keywords: transmembrane protein

F:69-85/Domain: transmembrane #status predicted <TM>

Query Match 12.0%; Score 89.5; DB 2; Length 700;

Best Local Similarity 23.5%; Pred. No. 24;

Matches 36; Conservative 29; Mismatches 47; Indels 41; Gaps 7;

Qy 9 NGKMSSTIVSEEDFILPVYKGELEKGYQDFGWEISGPEGKK----- 50

Db 430 NDTKSSST---QDALIKEIQ-DLEKGRP-----ELSDLTHKKYSIINHESVISKLATVEK 480

Qy 51 ---DAGYVNL-SKDTF---IKPVFKTIEKKEENKPTFDVSKKKNPQVNHSQLNESH 103

Db 481 TKADQKYFAAMRSKDSILIEIKTSLKSNEL-----ILQKSDRLLOQKIGMLH 533

Qy 104 RKEDLREHSQKSDKTDVATVLDKNNISSK 136

Db 534 KQLDLSQNNERRLDSSKTEILKIIDLNNSTK 566

RESULT 30

T19006

ankyrin related protein C06C3.1 - *Caenorhabditis elegans*

N:Contains: myosin-light-chain-phosphatase (EC 3.1.3.53)

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Mar-2000

C:Accession: T19006; T22086

R:Berks, M.

submitted to the EMBL Data Library, August 1994

A:Reference number: Z19058

A:Accession: T19006

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1016 <WIL>

A:Cross-references: EMBL:Z36719; PIDN:CAA85318.1; GSPDB:GN000020; CESP:C06C3.1

A:Experimental source: clone C06C3

R:Matthews, P.

submitted to the EMBL Data Library, January 1995

A:Reference number: Z19510

A:Accession: T22086

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1016 <WT2>

A:Cross-references: EMBL:Z47809; PIDN:CAA87782.1; GSPDB:GN000020; CESP:C06C3.1

A:Experimental source: clone F42A8

C:Genetics:

A:Gene: CESP:C06C3.1

A:Map position: 2

A:Introns: 27/3; 94/3; 279/3; 352/2; 462/2; 523/3; 569/2; 718/3; 766/3; 833/3; 89

C:Keywords: phosphoric monoester hydrolase

Query Match 12.0%; Score 89.5; DB 2; Length 1016;

Best Local Similarity 21.6%; Pred. No. 36;

Matches 33; Conservative 33; Mismatches 38; Indels 49; Gaps 8;

Qy 5 VTIONGKMSSTIVSEEDFILPVYKGELEKGYQDFGWEISGPEGKDGAGYVNLSKDTFI 64

Db 265 ILLENGAELSD-----LTTGADVLGVADKECIDYVELA-DTV- 302

Qy 65 KPVFKTIEEKKEE---NKPTFDVSKKKNPQVNHSQLNESHKREDIQ-REEHSQK--- 116

Db 303 -----KVQNKRSKPGSGSQPPTSILQEKVHRMPSHEEHVLTSEKRDLOHQDHSNEFL 357

Qy 117 -----SDSTKDVAT-VLDKNNISS 135

Db 358 HSHPSTASVSGTSSNTNTTTTIVIGENDISA 390

RESULT 31

E96795

unknown protein F28016.8 [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: E96795

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dewar, K.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: E96795

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-528 <STO>

A:Cross-references: GB:AB005173; NID:G6143888; PIDN:AAF04434.1; GSPDB:GN00141

C:Genetics:

A:Gene: F28016.8

A:Map position: 1

Query Match 11.9%; Score 89; DB 2; Length 528;

Best Local Similarity 24.3%; Pred. No. 19;

Matches 35; Conservative 22; Mismatches 51; Indels 36; Gaps 6;

Qy 31 ELEKGYQDFGW-----EISGPEGKKDAG-----YVNLKDTFTKPKVPKTBEE 73

Db 200 ELAYDYNFEWYGGAKVRCLCGAVACSGFLGAKSRGFOASPYVILIEDTY---VVEDGDD 256

Qy 74 KKEENKPTFDVSKK--DNPQVNHSQLNESHKREDLQREH-----SQKSPDKD- 122

Db 257 RYVDKIPVYDSAEDELTSPEKNGESNTNEEKEDISTENHLESTALNIQQSDSTPT 316

Qy 123 -----VTATVLDKNNISSKSTTN 141

Db 317 MEEDVTVETVKTETSEDWKLLSQN 340

RESULT 32

E89883

conserved hypothetical protein SA0976 [imported] - *Staphylococcus aureus* (strain N315)

C:Species: *Staphylococcus aureus*

C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: E89883

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc,

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: E89883

```

submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid C35E7.
A;Reference number: Z21278
A;Accession: T33068
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-301 <GRA>
A;Cross-references: EMBL:AF067216; PIDN:AAC17524.1; GSPDB:GN00019; CBSP:C35E7.9
A;Experimental source: strain Bristol N2; clone C35E7
C;Genetics:
A;Gene: CBSP:C35E7.9
A;Map position: 1
A;Introns: 30/3; 193/1; 236/2

Query Match          11.8%; Score 88.5; DB 2; Length 301;
Best Local Similarity 24.8%; Pred. No.11;
Matches 27; Conservative 17; Mismatches 38; Indels 27; Gaps 3;

Qy 35 GYQFDGWEISGPEGKK-----DAGVVLNSKDTIKVPVKKI-----EE 73
    ||:|||||
Db 10 GTIAGWILAGCGKKKDGKSTASAAAPKADSKMKPPVENVKKKSEKKEPKKEEP 69
    ||:|||||

Qy 74 KKEEENKPTFDYKSKDNQVNHSQLNSHRKEDLQREHSOKSDSTKD 122
    ||:|||||
Db 70 KKEEEKKESKKSEKKDKK-----EAKKEDDKKDKKEDKEDKKD 112
    ||:|||||

RESULT 35
A71683
hypothetical protein RP278 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C;Accession: A71683
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alismark, L
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499; PMID:9823893
A;Accession: A71683
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-371 <AND>
A;Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14739.1; PID:g3868083
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: RP278

Query Match          11.8%; Score 88.5; DB 2; Length 371;
Best Local Similarity 23.1%; Pred. No.14;
Matches 34; Conservative 33; Mismatches 37; Indels 43; Gaps 9;

Qy 11 KEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGPEGKK-DAGVVLNSKDTIKVPVK 69
    ||:|||||
Db 192 EQUNTKTIINEE-----FRNLQ-----EKKIEGKKTQASEALNK-----LKPIYK 234
    ||:|||||

Qy 70 KIEEKKEE-----ENKPTFDVSKKDNQVNV---HSQLNSHRKEDLQREHSOKSDSTK 121
    ||:|||||
Db 235 QMDEEREELFKLAELLPQY-----AQANIDKHAQVAKQVQTKIENDPNVKELEKIQ 286
    ||:|||||

Qy 122 DVTATVLDKNNISSKSTT-----NNPN 143
    ||:|||||
Db 287 E-----IVSKIYSKSKSTNDIIINPN 309
    ||:|||||

RESULT 36
G88436
protein T04A8.13 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: G88436
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916

```



C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Dec-2000  
C:Accession: A41859  
R:Poulsen, K.; Reinholdt, J.; Kilian, M.  
J. Bacteriol. 174, 2913-2921, 1992  
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty  
A:Reference number: A41859, MUID:92234949, PMID:1373717  
A:Accession: A41859  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1702 <POU>  
A:Cross-references: GB:M87489; NID:G148906; PIDN:AAA24966.1; PID:G148907  
A:Experimental source: strain HK715  
A>Note: sequence extracted from NCBI backbone (NCBIP:97282)  
C:Superfamily: IgA-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase

Query Match 11.7%; Score 87.5; DB 2; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 91;  
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;  
QY 56 INLSKDTFIKVPFKIEKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHS- 114  
Db 1296 INTGSATAITETAERKSDKPTETAASTDASQHKANTVADNSVANNSESSEPKRRRSI 1355  
QY 115 --QKSDSTKDVTVLDKNNISSKSTNNPNK 144  
Db 1356 SQPQETSAEETTAASTDETTIADNSKRKPNR 1387

Search completed: February 10, 2004, 10:58:33  
Job time : 12.3976 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 32.7596 Seconds  
(without alignments)  
697.707 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTVTIQNGKMSSTIVSE.....ATVLKKNISKSSTNNPNK 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.\*

```
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	748	100.0	773	22	AA198343
2	748	100.0	2120	21	AA1981710
3	748	100.0	2140	24	ABU01020
4	615	82.2	117	19	AAW55096
5	615	82.2	117	23	ABP54590
6	110	14.7	746	22	AA1981779
7	110	14.7	778	23	ABP39023
8	106.5	14.2	484	21	AA1984777
9	104	13.9	665	21	AA198278

10	101.5	13.6	564	22	AB19177	Drosophila melanog
11	92.5	12.4	2519	23	ABG16636	Novel human diagno
12	91	12.2	2060	23	AAE20967	Staphylococcus lug
13	89.5	12.0	511	20	AA135091	Chlamydia pneumoni
14	89.5	12.0	1408	22	AB191704	Drosophila melanog
15	89	11.9	645	24	AB191106	Pathogen specific
16	89	11.9	654	24	ABP56879	Staphylococcus epi
17	88.5	11.8	281	22	AB191704	Drosophila melanog
18	88.5	11.8	1141	22	AA191704	Shrimp white spot
19	88	11.8	225	23	ABP73992	Candida albicans e
20	88	11.8	258	22	AA191704	Human protein sequ
21	88	11.8	645	24	AB191704	Pathogen specific
22	88	11.8	817	22	AA191704	Human protein SEQ
23	88	11.8	817	22	AA191704	Human protein SEQ
24	87	11.6	607	22	AA191704	Human polypeptide
25	87	11.6	635	21	AA191704	Plasmodium falcipa
26	87	11.6	902	23	AB191704	Staphylococcus epi
27	86.5	11.6	1183	22	AB191704	Drosophila melanog
28	86.5	11.6	2485	21	AA191704	Plasmodium falcipa
29	86	11.5	209	21	AA191704	Human ORFX ORF2661
30	86	11.5	408	21	AA191704	Arabidopsis thalia
31	86	11.5	456	21	AA191704	Arabidopsis thalia
32	86	11.5	476	21	AA191704	Arabidopsis thalia
33	86	11.5	757	21	AA191704	Human cell cycle r
34	86	11.5	758	21	AA191704	Human colon cancer
35	86	11.5	922	22	AA191704	Shrimp white spot
36	86	11.5	1315	20	AA191704	S. aureus SdrD pro
37	86	11.5	1315	24	AB191704	Pathogen specific
38	86	11.5	1349	22	AA191704	Staphylococcus aur
39	86	11.5	1349	22	AA191704	Staphylococcus aur
40	86	11.5	1980	21	AA191704	Human sodium chann
41	86	11.5	1980	23	AA191704	Human sodium chann
42	85.5	11.4	402	20	AA191704	S. aureus trigger
43	85.5	11.4	434	20	AA191704	Staphylococcus hae
44	85.5	11.4	525	20	AA191704	S. aureus trigger
45	85.5	11.4	884	23	AB191704	Candida albicans e

#### ALIGNMENTS

#### RESULT 1

AA198343  
ID AA198343 standard; Protein; 773 AA.  
AC AA198343;  
XX  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE S. pneumoniae Spi30 polypeptide.  
KW Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;  
KW bronchial; lung; blood; infection; immune response; immunotherapy;  
KW antibacterial; auditory; vaccine.  
XX  
OS Streptococcus pneumoniae.  
PN WO200076540-A2.  
XX  
PD 21-DEC-2000.  
XX  
PF 09-JUN-2000; 2000WO-US15925.  
XX  
PR 10-JUN-1999; 99US-0138453.  
XX  
PA (MEDI-) MED IMMUNE INC.  
XX  
PI Adamou JE, Choi GH;  
XX  
DR WPI; 2001-112197/12.  
XX  
DR N-PSDB; AAC84742.  
XX  
PT New vaccines comprising Spi28 or Spi30 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by  
 PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
 PT blood infections  
 XX  
 XX Claim 8; Page 51-54; 54pp; English.  
 XX  
 CC The invention relates to novel immunogenic polypeptides, Sp128 and Sp130  
 CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
 CC the treatment and prevention of pneumococcal infections, particularly  
 CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
 CC bronchial, lung or blood infections. The antigens are used as immunogenic  
 CC agents to stimulate an immune response. The antisera and antibodies may  
 CC also be used in diagnosing and treating pneumococcal infections.  
 CC Recombinant polypeptides serve as a mechanism for stimulating production  
 CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
 CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the S. pneumoniae Sp130 polypeptide.  
 XX  
 XX SQ Sequence 773 AA;

Query Match 100.0%; Score 748; DB 22; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-67;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSK 60  
 DB 630 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSK 689  
 QY 61 DTFIKPVFKIIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 690 DTFIKPVFKIIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 749  
 QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
 DB 750 KDVTATVLDKNNISSKSTNNPNK 773

RESULT 2  
 ID AAY81710  
 XX AAY81710 standard; Protein; 2120 AA.  
 AC AAY81710;  
 XX  
 XX 02-JUN-2000 (first entry)  
 DE Streptococcus pneumoniae protein sequence ID3.  
 XX  
 KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 XX WO200006738-A2.  
 PN  
 XX 10-FEB-2000.  
 PD  
 XX  
 XX 27-JUL-1999; 99WO-GB02452.  
 PF  
 XX 27-JUL-1998; 98GB-0016336.  
 PR  
 XX 19-MAR-1999; 99US-0125329.  
 XX  
 XX (MICR-) MICROBIAL TECHNIQS LTD.  
 PA  
 XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
 PI  
 XX WPI; 2000-195301/17.  
 DR  
 XX N-PSDB; AA291806.  
 DR  
 XX Streptococcal proteins and polynucleotides useful for diagnosis,  
 PT treatment and prophylaxis of bacterial infections  
 PT  
 XX

PS Claim 2; Page 41-42; 76pp; English.  
 XX  
 CC This sequence represents a Streptococcus pneumoniae protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful as vaccines and also in  
 CC diagnostic assays. The sequences are useful for the detection or  
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
 CC infection. As the sequences can be used to treat S. pneumoniae infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
 CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
 CC meningitis.  
 XX  
 XX SQ Sequence 2120 AA;

Query Match 100.0%; Score 748; DB 21; Length 2120;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-67;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSK 60  
 DB 1943 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSK 2002  
 QY 61 DTFIKPVFKIIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 2003 DTFIKPVFKIIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 2062  
 QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
 DB 2063 KDVTATVLDKNNISSKSTNNPNK 2086

RESULT 3  
 ID ABU01020  
 XX ABU01020 standard; Protein; 2140 AA.  
 AC ABU01020;  
 XX  
 XX 11-FEB-2003 (first entry)  
 DE S. pneumoniae type 4 strain protein from coding region #590.  
 XX  
 KW Bacterial meningitis; pneumonia; sepsis; otitis media;  
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KW auditory; respiratory; gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae type 4 strain.  
 XX  
 XX WO200277021-A2.  
 PN  
 XX 03-OCT-2002.  
 PD  
 XX  
 XX 27-MAR-2002; 2002WO-IB02163.  
 PF  
 XX 27-MAR-2001; 2001GB-0007658.  
 PR  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX Masignani V, Tettelin H, Fraser C;  
 PI  
 XX WPI; 2003-040579/03.  
 DR  
 XX N-PSDB; ABX06302.  
 DR  
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
 PT

PT or ear infection -

PS Claim 1; SEQ ID No 1180; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS56454. Also included are an antibody which binds one of the  
 CC proteins, treating a patient by administering the protein, DNA or  
 CC antibody (in a composition), a kit comprising first and second primers,  
 CC which are the nucleic acid cited above or fragments between nucleotides  
 CC 8-100 of a sequence not defined in the specification, for amplifying a  
 CC target sequence contained within a Streptococcus nucleic acid sequence,  
 CC where the first primer is substantially complementary to the target  
 CC sequence and the second primer is substantially complementary to the  
 CC complement of the target sequence, and where the parts of the primers  
 CC having substantial complementarity define the termini of the target  
 CC sequence to be amplified, assay comprising contacting a test compound  
 CC with the protein, and determining whether the test compound binds to the  
 CC protein and a Streptococcus pneumoniae bacterium, where one or more  
 CC genes encoding the proteins has been rendered inactive. The proteins,  
 CC nucleic acid molecules, antibody and compositions are useful as  
 CC medicaments for treating or preventing a disease or infection due to  
 CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
 CC sepsis, otitis media or ear infection. They are also useful in developing  
 CC vaccines, diagnostics and antibiotics. The methods are useful for  
 CC identifying immunodominant proteins. The present sequence is one of  
 CC the 2469 proteins expressed by the identified coding regions from the  
 CC genomic sequence.

CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2140 AA;

Query Match 100.0%; Score 748; DB 24; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-67;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 60  
 DB 1963 HRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 2022  
 QY 61 DTFIKPVFKTIEKKKEENKPTFDVSKKKDNPQVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 2023 DTFIKPVFKTIEKKKEENKPTFDVSKKKDNPQVNHSQLNESHKEDLQREHSQKSDST 2082  
 QY 121 KDVTATVLDKKNISSKSTNNPNK 144  
 DB 2083 KDVTATVLDKKNISSKSTNNPNK 2106

RESULT 4

AAW55096

ID AAW55096 standard; Protein; 117 AA.

XX AC AAW55096;

DT 02-OCT-1998 (first entry)

XX Streptococcus pneumoniae SP0043 protein.

XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 KW detection; pneumonia; otitis media; meningitis.

XX Streptococcus pneumoniae.

XX WO9818930-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19422.  
 XX 31-OCT-1996; 96US-0029960.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 PA Choi GH, Hromockyj A, Johnson LS, Kunsch CA;  
 PI WPI; 1998-272224/24.  
 XX N-PSDB; AAV27357.

XX Nucleic acid encoding antigenic peptide(s) from Streptococcus  
 PT pneumoniae - or their epitope-containing fragments, useful in  
 PT protective or therapeutic vaccines, and for diagnosis  
 XX Claim 11; Page 62; 118pp; English.

XX The present sequence represents a protein from Streptococcus pneumoniae.  
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
 CC can be useful in vaccines for inducing protective antibodies against  
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 CC are used to detect Streptococcus infection (by usual hybridisation or  
 CC amplification methods), also for isolating Streptococcus genes or their  
 CC allelic variants. The protein can be used similarly to detect specific  
 CC antibodies in standard immunoassays, especially for diagnosing or  
 CC monitoring infections. Antibodies which bind the protein are used to  
 CC detect corresponding antigens, to purify the protein and for passive  
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
 CC (especially 10-300) mu g/ml per dose.

XX SQ Sequence 117 AA;

Query Match 82.2%; Score 615; DB 19; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 5e-55;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 YKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFKPVFKTIEKKKEENKPTFDVSK 87

DB 1 YKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFKPVFKTIEKKKEENKPTFDVSK 60

QY 88 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTATVLDKKNISSKSTNNPNK 144

DB 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTATVLDKKNISSKSTNNPNK 117

RESULT 5

ABP54590

ID ABP54590 standard; Protein; 117 AA.

XX AC ABP54590;

DT 04-SEP-2002 (first entry)

XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
 KW antibacterial; Streptococcal infection; detection.

XX Streptococcus pneumoniae.

XX US2002061545-A1.

XX 23-MAY-2002.

XX 22-JAN-2001; 2001US-0765272.

XX 30-OCT-1997; 97US-0961083.

XX (CHOI/) CHOI G H.

XX (KUNS/) KUNSCH C A.

PA (BARA/) BARASH S C.  
 PA (DILL/) DILLON P J.  
 PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX  
 DR WPI; 2002-479261/51.  
 DR N-PSDB; ABQ84825.  
 XX  
 XX New Streptococcus pneumoniae antigens, useful for detecting  
 PT Streptococcus and for preventing or attenuating disease caused by  
 PT Streptococcus infection -  
 XX  
 PS Claim 11; Page 29; 70pp; English.  
 XX  
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
 CC The S. pneumoniae antigens have antibacterial activity and can be  
 CC used in vaccines. The S. pneumoniae antigens can also be used to  
 CC prevent or attenuate a Streptococcus infection in an animal. The  
 CC polynucleotides encoding the S. pneumoniae antigens can be used to  
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
 CC which are used in an example from the present invention.  
 XX  
 SQ Sequence 117 AA;  
 Query Match 82.2%; Score 615; DB 23; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 5e-55;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 28 YKGELEKGYQFDGWETSGFEGKDDAGYVNLKDTFIKPVFKKIEEKNKPTFDVSK 87  
 Db 1 YKGELEKGYQFDGWETSGFEGKDDAGYVNLKDTFIKPVFKKIEEKNKPTFDVSK 60  
 QY 88 KKDNPQVNSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 144  
 Db 61 KKDNPQVNSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117  
 RESULT 6  
 AAG81779  
 ID AAG81779 standard; Protein; 746 AA.  
 XX  
 AC AAG81779;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.  
 XX  
 KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;  
 XX vaccination; endocarditis.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN WO200134809-A2.  
 XX  
 XX 17-MAY-2001.  
 PD  
 XX  
 XX 09-NOV-2000; 2000WO-US30782.  
 PF  
 XX  
 XX 09-NOV-1999; 99US-0164258.  
 PR  
 XX (GLAX ) GLAXO GROUP LTD.  
 PA  
 XX Kimmerly WJ;  
 PI  
 XX WPI; 2001-316495/33.  
 DR N-PSDB; AAH52629.  
 XX

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 PS Claim 18; Page 208; 218pp; English.  
 XX  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX  
 SQ Sequence 746 AA;  
 Query Match 14.7%; Score 110; DB 22; Length 746;  
 Best Local Similarity 27.0%; Pred. No. 0.027;  
 Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;  
 QY 5 VTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW---EISGFE-----G 48  
 Db 596 ITIGNGKQIKQOSVKSGTKVLPHSKVLMLTGELTMP-DMTGWTKEEDVLAFEDLTCLKVVS 654  
 QY 49 KKDAGYVNL--LSKDTFIKPVFKKIEEKNKPTFDVS-----KKDNPQVNSQLNES 102  
 Db 655 TKGNGFVTVNQISKGQIIK-----NKKIEVLSAEDTDDQKTDDESDN 701  
 QY 103 HRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141  
 Db 702 KSKDKADEHSHNTSSSTKN-----DKSNADSKNDSDD 734  
 RESULT 7  
 ABP39023  
 ID ABP39023 standard; Protein; 778 AA.  
 XX  
 AC ABP39023;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.  
 XX  
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 XX antibacterial; gene therapy.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 XX 30-APR-2002.  
 PD  
 XX  
 XX 13-AUG-1998; 98US-0134001.  
 PF  
 XX  
 XX 14-AUG-1997; 97US-055779P.  
 PR  
 XX 08-NOV-1997; 97US-064964P.  
 PR  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA  
 XX Doucette-Stamm LA, Bush D;  
 PI  
 XX



PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	08-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG	

PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection -  
XX  
XX  
XX Disclosure; Page 321-322; 577pp; English.  
XX  
CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (I) (especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX  
XX Sequence 665 AA;

Query Match 13.9%; Score 104; DB 21; Length 665;  
Best Local Similarity 25.9%; Pred. No. 0.093;  
Matches 42; Conservative 29; Mismatches 35; Indels 56; Gaps 10;  
Qy 1 HRVT-VTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEI--SGFEGKKDAGYVIN 57  
Db 138 HQNELNLQSGK-----NQDI-----NKNKKGKQ-----DISNSNAENKGD----- 174  
Qy 58 LSKDTFIKVPFKIEEKE-----EENKPTFD-----VSKKDNPNVHNSQLNESH 103  
Db 175 -----VKEGVKELEKEKKEKISDDHKVNEKKSDDHKVNEKKSDDHKVNEKKSDDH 228  
Qy 104 RKEDLQR--EHSQKSDSTKQVATVLDKNNISSTNNPNK 144  
Db 229 KIEVKVKEHEEBEE-----DKKEKSENKKNKDNK 261

RESULT 10  
ABB61977  
ID ABB61977 standard; Protein; 564 AA.  
XX  
XX ABB61977;  
XX  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 12723.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.  
DR N-PSDB; ABL06080.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 564 AA;

Query Match 13.6%; Score 101.5; DB 22; Length 564;  
Best Local Similarity 24.5%; Pred. No. 0.13;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
Qy 21 EDFILPVYKGELEKGYQFDGW-----EISGFEGKKDAGYVI-----NLSKDTFIK 65  
Db 78 EDLDTPLSESRFSK--VFDGWVDEHRDEHDGHDVQEPFSGEALDDHDDHDDHDEDEE 135  
Qy 66 PVFKKIEEKEENKPT-----FDVSKKDNPNVHNSQLNESHKEDLQREHHSQKSDS 119  
Db 136 PLTEEELEEEEEEPTTEDEPADEYEEDEENNA--GENITAEADAEEDND 193  
Qy 120 TKDVTATVLDKNNISST 138  
Db 194 EGTVEATVEATTEATTEAT 212

RESULT 11  
ABG16636  
ID ABG16636 standard; Protein; 2519 AA.  
XX  
XX ABG16636;  
XX  
XX 18-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #16627.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS80823.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in





CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX  
SQ Sequence 511 AA;  
Query Match 12.0%; Score 89.5; DB 20; Length 511;  
Best Local Similarity 24.5%; Pred. No. 2;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
Oy 64 IKPVFKIEKKEENKPTFD-----VSKKONQVNHSHK 105  
Db 95 VKGVFKTPQARPEVSSPRLPSHVHQRLGLEGFRDRIQRSENPEADLGKMKRSYSD 154  
Oy 106 EDLQREHSQKSDTKDVTATVLKNNISKSTT 139  
Db 155 GDLRVGHDSNEDSTEDSR-----EGGEPSSKSS 185

RESULT 14  
ABBS8704  
ID ABB58704 standard; Protein; 1408 AA.  
XX  
AC ABB58704;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 2904.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL02807.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 2904; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1408 AA;  
Query Match 12.0%; Score 89.5; DB 22; Length 1408;  
Best Local Similarity 24.7%; Pred. No. 8;  
Matches 48; Conservative 23; Mismatches 62; Indels 61; Gaps 9;  
Oy 9 NGKE---MSSTIVSEEDF--ILPVYKGELEKGYQFDGWEISGPEGK--KDAGYV----- 55  
Db 646 NGKQKLNAKSTKSEKDFEIVSVKKE-EKNDNQSEKLENDQDKLPKDDYLDVKPP 704  
Oy 56 -----INLSKDTFTKVPFK-----KIEKKKEENKPTFDVSKKONQVNHSHK 100  
Db 705 MESKEYSEPLAVKLGEDVSKSPSNESDTKISEVKGE--PKMNGELPKTGEQVKQPEKS 762  
Oy 101 ESHRK-EDLQREHSQKSDTKDVT-----ATVLDK 130  
Db 763 KTOAKIEFLAKVOSELKSKVDVTPKEQTKEETHKGFSKREATPKSEPKGTTEVDK 822  
Oy 131 NNISSKSTNNPNK 144  
Db 823 SNDVSVIKPSEPDE 836  
RESULT 15  
ABJ19106  
ID ABJ19106 standard; Protein; 645 AA.  
XX  
AC ABJ19106;  
XX  
DT 06-MAR-2003 (first entry)  
XX  
DE Pathogen specific antigen related staphylococcal protein SEQ ID No 414.  
XX  
KW Antibacterial; virucide; fungicide; protozoicide; cytostatic; anti-HIV;  
KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
KW autoimmune disease; HIV; hepatitis.  
XX  
OS Staphylococcus sp.  
XX  
PN WO200259148-A2.  
XX  
PD 01-AUG-2002.  
XX  
PF 21-JAN-2002; 2002WO-EP00546.  
XX  
PR 26-JAN-2001; 2001AT-0000130.  
XX  
PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.  
XX  
PI Meinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W;  
PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;  
PI Tempelmaier B;  
XX  
DR WPI; 2003-075410/07.  
XX  
XX Identifying, isolating and producing hyperimmune serum-reactive  
PT antigens from a pathogen, for preparing vaccine or medicament for  
PT treating or preventing e.g. staphylococcal infections, comprises  
PT providing antibody preparation -  
XX  
XX Example 7; Page 220; 252pp; English.  
XX  
CC The invention relates to a novel method for identifying, isolating and  
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
CC allergen, a tissue or host prone to auto-immunity, where the antigens  
CC are used in a vaccine, comprises providing antibody preparation from a  
CC plasma pool of a type of animal, or individual sera with antibodies  
CC against the specific pathogen, tumour, allergen, tissue or host prone to  
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of  
CC the 62 sequences of 53-2261 amino acids fully defined in the

specification, or their hyperimmune fragments are useful for the manufacture of a pharmaceutical preparation, particularly a vaccine against staphylococcal infections or colonisation against *S. aureus* or *S. epidermidis*. The preparation of antibodies is useful for the manufacture of a medicament for treating or preventing staphylococcal infections or colonisation against *S. aureus* or *S. epidermidis*. The antibody preparations may also be used for diagnostic and imaging purposes. Other conditions that can be treated include cancer, autoimmune diseases or infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or protozoan pathogens. This sequence represents a staphylococcal protein relating to the method for identifying and producing pathogen specific antigens of the invention.

```

Query Match      11.9%; Score 89; DB 24; Length 645;
Best Local Similarity 21.9%; Pred.No.3;
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;

Qy      6 .TIOGKEMSSITVSEEDFILPVYKGLE-KGYQF-----DGWRISGFEGKK----- 50
       :.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
Db      355 SVENNESMDTFVKH-----PIKTMLNGKKYMYWMTNDYDKDFVWGGRVPTISKDA 409
       :.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
Qy      51 -----DAGVINL-SKDTFIKPVFVKIIEKKKEE 78
       :.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
Db      410 KNNTRTIIFPVEGKTLDAIVKHVKTIDYGQYHVRIVDKEAFTKANTDKSNKEQQD 469
       :.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
Qy      79 NKPTFDV-----SKKNQPVNHSOLNESHREKDLO---REEHSQSDSTPKDYT-ATVL 128
       :.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
Db      470 NSAKKEATPATPTPSPEVESQKQDSQKDDNKQLPFSVEKENDASSESGKDTTPATKP 529
       :.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
Qy      129 DKNNISSKSTTNPNK 144
       :.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:
Db      530 TKGEVRSSTTT--PTK 543
       :.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:.:

```

RESULT 16	
ABP56879	
ID	ABP56879 standard; Protein; 654 AA.
XX	
AC	ABP56879;
XX	
DT	07-APR-2003 (first entry)
XX	
DE	Staphylococcus epidermidis KrKn protein SEQ ID NO:10.
XX	
KW	Staphylococcus aureus; Staphylococcus epidermidis; MSCRAMM; antibody;
KW	microbial surface component recognising adhesive matrix molecule;
KW	surface protein; infection; antibacterial; antiinflammatory; vaccine;
KW	immunosuppressive; antiarthritic; gene therapy; pneumonia; endocarditis;
KW	septic arthritis; biomaterial related infection.
XX	
OS	Staphylococcus epidermidis.
XX	
PN	WO2002102829-A2.
XX	
PD	27-DEC-2002.
XX	
PF	17-JUN-2002; 2002WO-US19220.
XX	
PR	15-JUN-2001; 2001US-298098P.
XX	
XX	(INH1-) INHIBITEX INC.
PA	(QUEB-) QUEEN ELIZABETH COLLEGE DUBLIN.
PA	(UYPA-) UNIV PAVIA.
XX	
PI	Foster TJ, Roche F, Patti JM, Hutchins JT, Hall A, Domanski P;
PI	Patel P, Syribeys P, Speziale P;
XX	
DR	WFI; 2003-167481/16.
DR	N-PSDB; ABZ22903.
XX	
PT	New isolated antibody that binds to a staphylococcal surface protein.

useful for treating or preventing *Staphylococcus aureus* infections,  
such as pneumonia, septic arthritis, endocarditis or biomaterial  
related infections -

Claim 1; Page 29-30; 122pp; English.

The present invention describes an isolated antibody (I) that binds to a  
staphylococcal surface protein selected from any of the 12 sequences of  
354-3283 amino acids given in ABP56875 to ABP56886. Also described:  
(1) an isolated antisera (II) comprising (I); (2) a diagnostic kit (III)  
comprising (I) and means for detecting binding by (I); (3) diagnosing  
(MI) an infection of *Staphylococcus aureus* by adding (I) to a sample  
suspected of being infected with the infection, and determining if  
antibodies have bound to the sample; (4) a pharmaceutical composition  
(IV) for treating or preventing an infection of *S. aureus* comprising (I),  
and a vehicle, carrier or excipient; (5) treating (M2) or preventing an  
infection of *S. aureus* by administering (I) to a human or animal patient;  
(6) producing (M3) an immunological response by administering to a human  
or animal an immunogenic amount of the isolated surface protein; (7) an  
isolated active fragment (V) from the A domain of the DsgA protein; and  
(8) a vaccine (VI) for treating or preventing an infection of *S. aureus*  
comprising the surface protein in an amount effective to elicit an immune  
response, and a vehicle, carrier or excipient. (I) has antiinflammatory,  
antibacterial, immunosuppressive and antiarthritic activities, and can be  
used in gene therapy. The antibody, composition and vaccine are useful  
for treating or preventing *Staphylococcus aureus* infection in a human or  
animal, such as pneumonia, septic arthritis, endocarditis or biomaterial  
related infections. The present sequence represents *Staphylococcus*  
*epidermidis* KrxN protein, which is used in the exemplification of  
the present invention.

Query Match	11.9%;	Score 89;	DB 24;	Length 654;
Best Local Similarity	21.9%;	Pred. No. 3.1;		
Matches	43;	Conservative	26;	Mismatches 63; Indels 64; Gaps 9;
Qy	6	TIQNGEMSS	TIYSEEDFILPVYKGLE-KGYQF-----	DGWEISGFEKK-----50
Db	364	SVENNESMD	TFVKH-----PIKTGMLNGKYYVMET	TNDYKDFWVEGQVRVTSKDA418
Qy	51	-----	-----	DAGYVNL-SKDTFFKVPFKIEKKEEE78
Db	419	KNNTRTIIF	PVYSEGKTLDAIVKVHVKTIDYDQGYHVR	IVDKAFTKANTDKSNKKEQOD478
Qy	79	NKPTFDV----	SKKXDNPPWNHSHRKEDLQ-----	REEHSQKSDSTKQVATVL128
Db	479	NSAKKATPATP	SKPTFPSPVEKESQKDSQKDDNKQLPSV	EKENDASESQKDTPTATPK538
Qy	129	DKNNISKST	TNNPNX144	
Db	539	TKGEVSSST	--PTK552	

RESULT 17	
ABB64828	
ID	ABB64828 standard; Protein; 281 AA.
XX	
XX	ABB64828;
AC	
XX	
XX	26-MAR-2002 (first entry)
DT	
DT	
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 21276.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL08931.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 21276; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 281 AA;  
 Query Match 11.8%; Score 88.5; DB 22; Length 281;  
 Best Local Similarity 22.1%; Pred. No. 1.1;  
 Matches 31; Conservative 34; Mismatches 52; Indels 23; Gaps 5;  
 QY 9 NGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVI-----NLS 59  
 DB 137 NFEFTNMTTQDLDLKNY-SLLDK-----DNEGAIITSKLGWVIRALGRQPNESIA 188  
 QY 60 KDTFKVPFKKIEE-KKEEENKPTFDVSKKDNQVNHSQLNE-----SHRKEDLQREH 113  
 DB 189 KEFCNVLIRKMHDTNKEELRDAFRVDFKKNNGYISTTELRAVFMALGKLEDDLEEM 248  
 QY 114 SOKSDSTKDVATVLDKNNI 133  
 DB 249 IREYLDQDNHINFESNNM 268  
 RESULT 18  
 AAG85008  
 ID AAG85008 standard; Protein; 1141 AA.  
 AC AAG85008;  
 XX  
 DT 11-SEP-2001 (first entry)  
 DE Shrimp white spot Bacilliform virus (WSBV) protein 99.  
 XX  
 KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;  
 KW antiviral agent; gene expression; antisense construct;  
 KW transgenic viral resistant shrimp.  
 XX  
 OS White spot syndrome virus.  
 XX  
 PN WO200138351-A2.  
 XX  
 PD 31-MAY-2001.  
 XX  
 PF 08-NOV-2000; 2000WO-US28888.  
 XX  
 PR 24-NOV-1999; 99CN-0124717.  
 XX

PA (PENY-) PE CORP NY.  
 PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.  
 PA (SINO-) SINOGENOMAX CO LTD.  
 XX  
 PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;  
 XX  
 DR WPI; 2001-355877/37.  
 DR N-PSDB; AAH62788.  
 XX  
 PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus  
 PT (WSBV), useful for producing viral polypeptides that can be used to  
 PT screen for agents that are useful for treating WSBV infection -  
 XX  
 PS Claim 1; Figure 3; 626pp; English.  
 XX  
 CC The invention provides the primary nucleotide sequence of the WSBV genome  
 CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and  
 CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences  
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid  
 CC molecules and proteins of the invention are useful for diagnosis and  
 CC monitoring viral infection, in screens for antiviral agents and for  
 CC monitoring viral gene expression or activity during a treatment regimen.  
 CC The nucleic acid molecules are also useful as antisense constructs to  
 CC control viral gene expression in infected cells and tissues and to create  
 CC transgenic viral resistant shrimp.  
 XX  
 SQ Sequence 1141 AA;  
 Query Match 11.8%; Score 88.5; DB 22; Length 1141;  
 Best Local Similarity 23.3%; Pred. No. 7.5;  
 Matches 30; Conservative 31; Mismatches 43; Indels 25; Gaps 4;  
 QY 15 STIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKSDTFIKPVFKKIE 72  
 DB 621 SNVEEEE-----EEEQMEEEEEEEVEEESGDKEDGDA-----PAQEEME 662  
 QY 73-EKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNN 132  
 DB 663 EEKEEQ-----QQPEESNGNQEEREEQQQQQPEREENKQADSDSDSSSSSSS 717  
 QY 133 ISSKSTNN 141  
 DB 718 SSSSSSSSS 726  
 RESULT 19  
 ABP73992  
 ID ABP73992 standard; Protein; 225 AA.  
 AC ABP73992;  
 XX  
 DT 30-JAN-2003 (first entry)  
 DE Candida albicans essential protein SEQ ID NO 7829.  
 XX  
 KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;  
 KW signal transduction; DNA replication; cell division; growth;  
 KW proliferation; Candida albicans; fungicide; antifungal.  
 XX  
 OS Candida albicans.  
 XX  
 PN WO200253728-A2.  
 XX  
 PD 11-JUL-2002.  
 XX  
 PF 26-DEC-2001; 2001WO-US49486.  
 XX  
 PR 29-DEC-2000; 2000US-259128P.  
 PR 20-FEB-2001; 2001US-0792024.  
 PR 22-AUG-2001; 2001US-314050P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX

PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
 XX WPI; 2002-566694/60.  
 DR N-PSDB; AB232542.  
 XX  
 PT Constructing strains for identifying gene products as effective targets  
 PT for therapeutic intervention, by inactivating in the strain one allele  
 PT of a gene and placing other allele of the gene under conditional  
 PT expression -  
 XX  
 XX Claim 44; SEQ ID NO 7829; 167pp + Sequence Listing; English.  
 XX  
 CC The invention relates to constructing (M1) a strain of diploid fungal  
 CC cells in which both alleles of a gene are modified, comprising modifying  
 CC one allele by insertion or replacement by a cassette having an  
 CC expressible selectable marker and modifying other allele by  
 CC recombination, of a promoter replacement fragment of a heterologous  
 CC promoter, so that expression of the second allele is regulated by the  
 CC promoter. (M1) is useful for constructing a strain of diploid fungal  
 CC cells in which both alleles of a gene are modified. The diploid fungal  
 CC cells having both alleles modified are useful for identifying a gene that  
 CC is essential to the survival or growth of a fungus, a gene that  
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
 CC that contributes to the resistance of a diploid fungus to an antifungal  
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
 CC and for identifying a therapeutic agent for treatment of a mammalian  
 CC disease. (M1) is useful for identifying a compound which modulates the  
 CC activity of a gene product, preferably enzymatic activity, carbon  
 CC compound catabolism, biosynthetic, transporter, transcriptional,  
 CC translational, signal transduction, DNA replication and cell division  
 CC activity. The method is useful for identifying a compound having the  
 CC ability to inhibit growth or proliferation of C. albicans cells and for  
 CC treating infection by C. albicans. The present sequence is that of an  
 CC essential Candida albicans protein used in the method of the invention.  
 CC Note: The sequence data for this patent is not represented in the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office.  
 XX  
 SQ Sequence 225 AA;  
 Query Match 11.8%; Score 88; DB 23; Length 225;  
 Best Local Similarity 28.0%; Pred. No. 0.89; Mismatches 43; Indels 12; Gaps 4;  
 Matches 30; Conservative 22;  
 QY 36 YQFDGWEISGFEKKDAGYVNLKDTFIKVPFKK----IEKKKEENKPTFDVSKK-- 89  
 DB 80 YDDDDDEFEFGSSNGAAKELNLSQAIEKWKQRDLIEIEKLNKSKKEIEIKAKS 139  
 QY 90 --DNQVNSQLNESHKEDLQREEH--SQKSDSTKDVTTATVLDKNN 132  
 DB 140 TIDDFYENYSNRDNRHOKELSEIQEKFIKRDFFLK--RGTLDWRVN 184  
 RESULT 20  
 AAB94584  
 ID AAB94584 standard; Protein; 258 AA.  
 XX  
 AC AAB94584;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:15383.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX

PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 XX  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 XX Claim 8; SEQ ID 15383; 2537pp + CD ROM; English.  
 PS  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 258 AA;  
 Query Match 11.8%; Score 88; DB 22; Length 258;  
 Best Local Similarity 26.3%; Pred. No. 1.1;  
 Matches 31; Conservative 25; Mismatches 50; Indels 12; Gaps 4;  
 QY 28 YKGELEKGYQFDGWEISGFEKKDAGY--VNLKDTFIKVPFKKIEKKKEENKPTFDV 85  
 DB 39 WEGEDEDEDVKNWDDDDDDKEEAEVKEPEVKESEK---KKAIEKIEKEKQKKQKEEI 95  
 QY 86 SKKKNPQ-----VNHSQLNESHKEDLQREHSHSQKSDSTKDVTTATV--LDKNNISSK 136  
 DB 96 KKRLEPEPKVLTPEEQQLADKLRLKQLQESDLELAKETFGVNTVYIDAMNPSR 153  
 RESULT 21  
 ABJ18979  
 ID ABJ18979 standard; Protein; 645 AA.  
 XX  
 AC ABJ18979;  
 XX  
 DT 06-MAR-2003 (first entry)  
 XX  
 DE Pathogen specific antigen related staphylococcal protein SEQ ID No 142.  
 XX  
 KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;  
 KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
 KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
 KW autoimmune disease; HIV; hepatitis.  
 XX

```

OS Staphylococcus sp.
XX WO200259148-A2.
XX 01-AUG-2002.
XX 21-JAN-2002; 2002WO-EP00546.
XX 26-JAN-2001; 2001AT-0000130.
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX Meinke A, Nagy E, Von Ahsen U, Klade C, Henics T, Zauner W;
XX Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;
XX Tempelmaier B;
XX WPI; 2003-075410/07.
XX Identifying, isolating and producing hyperimmune serum-reactive
XX antigens from a pathogen, for preparing vaccine or medicament for
XX treating or preventing e.g. staphylococcal infections, comprises
XX providing antibody preparation -
XX Claim 21; Page 168; 252pp; English.
XX The invention relates to a novel method for identifying, isolating and
XX producing hyperimmune serum-reactive antigens from a pathogen, tumour,
XX allergen, a tissue or host prone to auto-immunity, where the antigens
XX are used in a vaccine, comprises providing antibody preparation from a
XX plasma pool of a type of animal, or individual sera with antibodies
XX against the specific pathogen, tumour, allergen, tissue or host prone to
XX auto-immunity. The hyperimmune serum-reactive antigens comprising any of
XX the 62 sequences of 53-2261 amino acids fully defined in the
XX specification, or their hyperimmune fragments are useful for the
XX manufacture of a pharmaceutical preparation, particularly a vaccine
XX against staphylococcal infections or colonisation against S. aureus or S.
XX epidermidis. The preparation of antibodies is useful for the manufacture
XX of a medicament for treating or preventing staphylococcal infections or
XX colonisation against S. aureus or S. epidermidis. The antibody
XX preparations may also be used for diagnostic and imaging purposes. Other
XX conditions that can be treated include cancer, autoimmune diseases or
XX infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
XX protozoan pathogens. This sequence represents a staphylococcal protein
XX relating to the method for identifying and producing pathogen specific
XX antigens of the invention.
XX Sequence 645 AA;
XX
XX Query Match 11.8%; Score 88; DB 24; Length 645;
XX Best Local Similarity 21.9%; Pred. No. 3.8;
XX Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;
XX
XX QY 6 TIQKEMSSTVSEEDFILPVYKGLB-KGYQF-----DGWEISGPEGKK----- 50
XX DB 355 SVNNESMMDTFVKH-----PIKTGLMGKKYMYMETTNDYDKDFMVEGQVRTISKDA 409
XX QY 51 -----DAGVINI--SKDTFIKPVFKKIEKKKEE 78
XX DB 410 KNRTTIIPVYEGTKLDAIVKHVHTIDYDGGVHVRVDKEAFTKANTDKSNKKEQD 469
XX QY 79 NKPTFDV-----SKKNDPQVNHSQLNESHKEDLQ-----REHSQKSDSTKDVLT-ATVL 128
XX DB 470 NSAKEATPATPSKPTSPVEKESQKQSDQKDNKQLPSVEKENDASSESGKVTLATKP 529
XX QY 129 DKNNISSKSTTNNPNK 144
XX DB 530 TKGEVSSSTT--PTK 543
XX
XX RESULT 22
XX AAM79318
XX ID AAM79318 standard; Protein; 817 AA.
XX
AC AAM79318;
XX 06-NOV-2001 (first entry)
XX Human protein SEQ ID NO 2964.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX N-PSDB; AAKS2451.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX Claim 20; Page 215; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX Sequence 817 AA;
XX
XX Query Match 11.8%; Score 88; DB 22; Length 817;
XX Best Local Similarity 27.3%; Pred. No. 5.3;
XX Matches 33; Conservative 21; Mismatches 55; Indels 12; Gaps 4;
XX
XX QY 30 GELEKGQFDGWEISG--FECKKDAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDYSK 87
XX DB 196 GQEKQKQSFKSWEASGKHQEVSPAVSLERQKQDTSKLRSTLPEEQKQKQKSPSPSQ 255
XX QY 88 -KDNPNVNHSQLNESHKED-----LQREHSQKSDSTKDVLTATVLDKNNISSKST 138
XX DB 256 WKQDTPKSKAGYVQEEHKKQETPKLWVQLQKEQ-DPKKQTPKSWTFSMQSEQNTTKSWT 314
XX QY 139 T 139

```

Db 315 T 315

RESULT 23  
AAM79319

ID AAM79319 standard; Protein; 817 AA.

XX  
AC AAM79319;

XX  
DT 06-NOV-2001 (first entry)

XX  
DE Human protein SEQ ID NO 2965.

XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.

XX  
OS Homo sapiens.

XX  
PN WO200157190-A2.

XX  
PD 09-AUG-2001.

XX  
PF 05-FEB-2001; 2001WO-US04098.

XX  
PR 03-FEB-2000; 2000US-0496914.

XX  
PR 27-APR-2000; 2000US-0560875.

XX  
PR 20-JUN-2000; 2000US-0598075.

XX  
PR 19-JUL-2000; 2000US-0620325.

XX  
PR 01-SEP-2000; 2000US-0654936.

XX  
PR 15-SEP-2000; 2000US-0663561.

XX  
PR 20-OCT-2000; 2000US-0693325.

XX  
PR 30-NOV-2000; 2000US-0728422.

XX  
PA (HYSE-) HYSEQ INC.

XX  
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;

XX  
PI WPI; 2001-476283/51.

XX  
DR N-PSDB; AAK52452.

XX  
PT Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -

XX  
PS Claim 20; Page 215; 6221pp; English.

XX  
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

XX  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

XX  
SQ Sequence 817 AA;

Query Match 11.8%; Score 88; DB 22; Length 817;  
Best Local Similarity 27.3%; Pred. No. 5.3;  
Matches 33; Conservative 21; Mismatches 55; Indels 12; Gaps 4;

QY 30 GELKGYQDGFWEISG--PEGKDGAVINLSKDTFIKVFKKIKREKKEENKPTFDVSK 87  
DB 196 GQFKKQESFSAEASGKHQEVSPAVSLQKQDTSLKRLSTLPEQKKQKQKESKSPSPSQ 255

QY 88 -KKONPQVNHQSOLNESHKED-----LQREHSQKSDSTKDYATATVLDKNNISSKST 138  
DB 256 WKQDTFKSGAGYQVEHKKQETPKLWPVQLQKEQ-DPKKQTPKSWTPSQSEQNTTKGWT 314

QY 139 T 139

DB 315 T 315

RESULT 24  
AAM39950

ID AAM39950 standard; Protein; 607 AA.

XX  
AC AAM39950;

XX  
DT 22-OCT-2001 (first entry)

XX  
DE Human polypeptide SEQ ID NO 3095.

XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.

XX  
OS Homo sapiens.

XX  
PN WO200153312-A1.

XX  
PD 26-JUL-2001.

XX  
PF 26-DEC-2000; 2000WO-US34263.

XX  
PR 21-JAN-2000; 2000US-0488725.

XX  
PR 25-APR-2000; 2000US-0552317.

XX  
PR 09-JUL-2000; 2000US-0598042.

XX  
PR 19-JUL-2000; 2000US-0620312.

XX  
PR 03-AUG-2000; 2000US-0653450.

XX  
PR 14-SEP-2000; 2000US-0662191.

XX  
PR 19-OCT-2000; 2000US-0693036.

XX  
PR 29-NOV-2000; 2000US-0727344.

XX  
PA (HYSE-) HYSEQ INC.

XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX  
PI WPI; 2001-442253/47.

XX  
DR N-PSDB; AAI59106.

XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -

XX  
PS Example 4; SEQ ID NO 3095; 10078pp; English.

XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM39642-AAK42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders

XX  
CC Note: The sequence data for this patent did not form part of the printed

```
CC specification.
XX
SQ Sequence 607 AA;
Query Match 11.6%; Score 87; DB 22; Length 607;
Best Local Similarity 33.3%; Pred. No. 4.5;
Matches 20; Conservative 12; Mismatches 28; Indels 0; Gaps 0;
QY 59 SKDTFIKPVFKKIEBKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSOKSD 118
Db 373 SPDNCCNELFKKKKKKKKKKKKKKKKKKKKKRRRRRRKKREKKREKKREKKREKD 432
RESULT 25
AAB18176
ID AAB18176 standard; Protein; 635 AA.
XX
AC AAB18176;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:33.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
FN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
XX (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX
DR
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 85-86; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are rifins or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasitic lifecycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
```

```
SQ Sequence 635 AA;
Query Match 11.6%; Score 87; DB 21; Length 635;
Best Local Similarity 23.8%; Pred. No. 4.8;
Matches 30; Conservative 20; Mismatches 52; Indels 24; Gaps 4;
QY 32 LEKGYQFDGWEISGEGKKDAGYVINLSKDTFFIKPVFKKIEBKKEENKPTFDVSKKDN 91
Db 202 LQKYNIQDDDEEDNETIRSDSKLRDIYDSQSKDIMSSSPNKEES-----NSSDNHN 256
QY 92 PQVNHSQLNESHKEDLQREHSOKSDSKDVTATVLD-----KNNISSKS 137
Db 257 KDINSS-----DNQNKDINSDDHN-NDSTNESITTSLSINNTNRNKKRKNININ 311
QY 138 TTNNPN 143
Db 312 NNNNSN 317
RESULT 26
ABP40312
ID ABP40312 standard; Protein; 902 AA.
XX
AC ABP40312;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5157.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
FN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
XX 08-NOV-1997; 97US-064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX WPI; 2002-381255/41.
XX N-PSDB; ABN92857.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS Disclosure; SEQ ID 5157; 267pp; English.
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences
XX can also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life
XX cycle or inhibit S. epidermidis infection.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO web site.
XX
SQ Sequence 902 AA;
Query Match 11.6%; Score 87; DB 23; Length 902;
Best Local Similarity 25.3%; Pred. No. 7.8;
Matches 38; Conservative 27; Mismatches 65; Indels 20; Gaps 7;
```







PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 08-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.

```
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.5%; Score 86; DB 21; Length 408;
Best Local Similarity 23.6%; Pred. No. 3.3;
Matches 35; Conservative 28; Mismatches 55; Indels 30; Gaps 7;

Qy 1 HRTVTIQNGEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGY--VINL 58
Db 209 HPVSAFLVYANERRAALREKNKSVVEVAK-----ITGEWNKLSDKKAPYKVA 260
Qy 59 SKDTFIKPVFKIEKEENKPTFDVSKKDNPOVNHSQLNESHKREDLQREHSQKSD 118
Db 261 NKEYILO-AMEYKRTKEE-----ALSQKE-----EEELLKLHQEALQWLKKEKTD 309
Qy 119 SKDQVATVLDKNISSKSTNN--PNK 144
Db 310 N-----LIKKEKATKKKKKNVDPNK 330

RESULT 31
AAG37133
ID AAG37133 standard; Protein; 456 AA.
XX AC AAG37133;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 45609.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PF 25-FEB-1999; 99US-0121825.
XX PF 05-MAR-1999; 99US-0123180.
XX PF 09-MAR-1999; 99US-0123548.
XX PF 23-MAR-1999; 99US-0125788.
XX PF 25-MAR-1999; 99US-0126264.
XX PF 29-MAR-1999; 99US-0126785.
XX PF 01-APR-1999; 99US-0127462.
XX PF 06-APR-1999; 99US-0128234.
XX PF 08-APR-1999; 99US-0128714.
XX PF 16-APR-1999; 99US-0129845.
XX PF 19-APR-1999; 99US-0130077.
XX PF 21-APR-1999; 99US-0130449.
XX PF 23-APR-1999; 99US-0130510.
XX PF 28-APR-1999; 99US-0130891.
XX PF 30-APR-1999; 99US-0131449.
XX PF 30-APR-1999; 99US-0132048.
XX PF 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
```



PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0137282.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 08-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.

```
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.5%; Score 86; DB 21; Length 476;
Best Local Similarity 23.6%; Pred. No. 4;
Matches 35; Conservative 28; Mismatches 55; Indels 30; Gaps 7;

QY 1 HRTVTIQKEMSTIVSEEDFILPVYKGELEKGVQFDGWEISGEGKKDAGY--VINL 58
Db 277 HPVSAPLVANERRAALRENSVVEVAK-----ITGEWKNLSDKKAPYEKVAKK 328

QY 59 SMDTIKPVFKIEBKKEENKPTFDVSKKNDPNVHNSQLNESHKEDLQREHSOKSD 118
Db 329 NKETVLO-AMEYKKTKEE-----ALSQKE-----EELKLHKQALQMLKKKKTID 377

QY 119 SPKQVATVLDKNNISKSTTN--PNK 144
Db 378 N-----LIKKEKATKKKKNNVDPNK 398

RESULT 33.
AAY44364
ID AAY44364 standard; protein; 757 AA.
AC AAY44364;
XX DT 14-MAR-2000 (first entry)
XX DE Human cell cycle regulation protein-5.
XX KW CECRP-5; cell cycle regulation protein-5; cell proliferation;
XX KW cell proliferative disease; cancer; atherosclerosis; cirrhosis;
XX KW hepatitis; psoriasis; immune system disorder; allergy; asthma;
XX KW acquired immune deficiency syndrome; Crohn's disease; Blast method;
XX KW rheumatoid arthritis; gene therapy; chromosomal mapping.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Modified-site 4
FT /note= "Potential glycosylation site"
FT Modified-site 63
FT /note= "Potential glycosylation site"
FT Modified-site 479
FT /note= "Potential glycosylation site"
FT Modified-site 498
FT /note= "Potential glycosylation site"
FT Modified-site 31
FT /note= "Potential phosphorylation site"
FT Modified-site 43
FT /note= "Potential phosphorylation site"
FT Modified-site 81
FT /note= "Potential phosphorylation site"
FT Modified-site 96
FT /note= "Potential phosphorylation site"
FT Modified-site 99
FT /note= "Potential phosphorylation site"

130 Modified-site /note= "Potential phosphorylation site"
158 Modified-site /note= "Potential phosphorylation site"
162 Modified-site /note= "Potential phosphorylation site"
163 Modified-site /note= "Potential phosphorylation site"
184 Modified-site /note= "Potential phosphorylation site"
186 Modified-site /note= "Potential phosphorylation site"
188 Modified-site /note= "Potential phosphorylation site"
271 Modified-site /note= "Potential phosphorylation site"
277 Modified-site /note= "Potential phosphorylation site"
287 Modified-site /note= "Potential phosphorylation site"
289 Modified-site /note= "Potential phosphorylation site"
323 Modified-site /note= "Potential phosphorylation site"
343 Modified-site /note= "Potential phosphorylation site"
367 Modified-site /note= "Potential phosphorylation site"
370 Modified-site /note= "Potential phosphorylation site"
432 Modified-site /note= "Potential phosphorylation site"
488 Modified-site /note= "Potential phosphorylation site"
499 Modified-site /note= "Potential phosphorylation site"
207 Modified-site /note= "Potential phosphorylation site"
277 Modified-site /note= "Potential phosphorylation site"
298 Modified-site /note= "Potential phosphorylation site"
370 Modified-site /note= "Potential phosphorylation site"
419 Modified-site /note= "Potential phosphorylation site"
474 Modified-site /note= "Potential phosphorylation site"
488 Modified-site /note= "Potential phosphorylation site"
499 Modified-site /note= "Potential phosphorylation site"
528 Modified-site /note= "Potential phosphorylation site"
532 Modified-site /note= "Potential phosphorylation site"
539 Modified-site /note= "Potential phosphorylation site"
609 Modified-site /note= "Potential phosphorylation site"
624 Modified-site /note= "Potential phosphorylation site"
730 Modified-site /note= "Potential phosphorylation site"

WO9964593-A2.
16-DEC-1999.
08-JUN-1999; 99WO-US12906.
08-JUN-1998; 98US-0088695.
XX PN
XX PD
XX XX
XX PF
XX XX
XX PR
XX XX
```

PA (INCY-) INCYTE PHARM INC.  
 XX Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR;  
 PI Patterson C;  
 XX WPI; 2000-105887/09.  
 DR N-PSDB; AAZ29484.  
 XX Novel regulatory proteins, for diagnosis, treatment and prevention of  
 PT cell proliferative and immune system diseases -  
 PT Claim 1; Page 70-71; 87pp; English.  
 XX The present sequence is cell cycle regulation protein-5 (CECRP-5). Blast  
 XX analytical method was used to identify this protein. CECRPs are  
 CC activators of cell proliferation or inhibitors of cellular processes that  
 CC modulate proliferation. They are used to treat or prevent cell  
 CC proliferative diseases like cancers, atherosclerosis, cirrhosis,  
 CC hepatitis, psoriasis, immune system disorders (e.g. acquired immune  
 CC deficiency syndrome, allergy, asthma, Crohn's disease, rheumatoid  
 CC arthritis). Antibodies are raised to screen for specific binding agents.  
 CC The corresponding nucleic acid is used in gene therapy, chromosomal  
 CC mapping and isolation of related sequences.  
 XX Sequence 757 AA;  
 SQ Query Match 11.5%; Score 86; DB 21; Length 757;  
 Best Local Similarity 29.1%; Pred. No. 7.7;  
 Matches 44; Conservative 13; Mismatches 52; Indels 42; Gaps 8;  
 QY 9 NGKMSSTIVSE--EDFILPVYKGELEKGYOF-----DGWE-----ISGFEKK 50  
 Db 409 HGKSDSPNVYTKKEIAILRLERLTLELKLTFEQQRSLWERYVEAKDQNGKQGTGKK 468  
 QY 51 DAG-----VVINLSKDTFKPVFKIEEKEENKPTFDVSKKKNPQVNH-----SQLNE 101  
 Db 469 KGGGRSHRAKNKSKETFLGSV-----KETFDMKNSTKEFVRHHKEIKQAKE 516  
 QY 102 SHRKEDLQREHSOKSD--STKQVTTATVLDK 130  
 Db 517 A-VKENLKKFSDSVKSTFRHFKDTTKNIFDE 546  
 RESULT 34  
 AAB53319  
 ID AAB53319 standard; Protein; 758 AA.  
 XX AAB53319;  
 AC  
 XX 09-MAR-2001 (first entry)  
 XX Human colon cancer antigen protein sequence SEQ ID NO:859.  
 DE Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;  
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;  
 KW neural disorder; immune system disorder; muscular disorder;  
 KW reproductive disorder; gastrointestinal disorder; renal disorder;  
 KW infectious disease; cardiovascular disorder.  
 XX Homo sapiens.  
 OS  
 XX WO200055351-A1.  
 PN  
 XX 21-SEP-2000.  
 PD  
 XX 08-MAR-2000; 2000WO-US05883.  
 PF  
 XX 12-MAR-1999; 99US-0124270.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX

PI Rosen CA, Ruben SM;  
 XX WPI; 2000-587534/55.  
 DR N-PSDB; AAC98076.  
 XX Colon cancer associated gene sequences, referred to as colon cancer  
 PT antigens, useful for the treatment, prevention, and diagnosis of colon  
 PT disorders such as colon cancer -  
 XX Claim 11; Page 1408-1411; 2104pp; English.  
 XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
 CC human colon cancer antigens can have cytosolic, cardioactive, muscular;  
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
 CC vulnary, nephrotropic, antiinfective and antibacterial activities, and  
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
 CC proteins and antibodies to the proteins are useful for the prevention,  
 CC treatment and diagnosis of colon disorders, such as colon cancer. The  
 CC polynucleotides may be used in diagnostics and research, such as for  
 CC chromosome identification, and as hybridisation probes. The proteins  
 CC may also be used to prevent diseases such as neural disorders, immune  
 CC system disorders, muscular disorders, reproductive disorders,  
 CC gastrointestinal disorders, wounds, renal disorders, infectious  
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
 CC AAB54007 represent sequences used in the exemplification of the present  
 CC invention.  
 XX Sequence 758 AA;  
 SQ Query Match 11.5%; Score 86; DB 21; Length 758;  
 Best Local Similarity 29.1%; Pred. No. 7.7;  
 Matches 44; Conservative 13; Mismatches 52; Indels 42; Gaps 8;  
 QY 9 NGKMSSTIVSE--EDFILPVYKGELEKGYOF-----DGWE-----ISGFEKK 50  
 Db 410 HGKSDSPNVYTKKEIAILRLERLTLELKLTFEQQRSLWERYVEAKDQNGKQGTGKK 469  
 QY 51 DAG-----VVINLSKDTFKPVFKIEEKEENKPTFDVSKKKNPQVNH-----SQLNE 101  
 Db 470 KGGGRSHRAKNKSKETFLGSV-----KETFDMKNSTKEFVRHHKEIKQAKE 517  
 QY 102 SHRKEDLQREHSOKSD--STKQVTTATVLDK 130  
 Db 518 A-VKENLKKFSDSVKSTFRHFKDTTKNIFDE 547  
 RESULT 35  
 AAG85023  
 ID AAG85023 standard; Protein; 922 AA.  
 XX AAG85023;  
 AC  
 XX 11-SEP-2001 (first entry)  
 DT Shrimp white spot Bacilliform virus (WSBV) protein 114.  
 DE Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;  
 XX antiviral agent; gene expression; antisense construct;  
 KW transgenic viral resistant shrimp.  
 KW  
 XX White spot syndrome virus.  
 OS  
 XX WO200138351-A2.  
 PN  
 XX 31-MAY-2001.  
 PD  
 XX 08-NOV-2000; 2000WO-US28888.  
 PF  
 XX 24-NOV-1999; 99CN-0124717.  
 PR  
 XX (PENY-) PE CORP NY.  
 PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.  
 PA

PA (SINO-) SINGENOMAX CO LTD.  
 XX Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;  
 XX WPI: 2001-355877/37.  
 DR N-PSDB; AAH62803.  
 XX Primary nucleotide sequence of the shrimp white spot bacilliform virus  
 PT (WSBV), useful for producing viral polypeptides that can be used to  
 PT screen for agents that are useful for treating WSBV infection -  
 XX Claim 1; Figure 3; 626pp; English.  
 XX The invention provides the primary nucleotide sequence of the WSBV genome  
 CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and  
 CC encoded proteins (AAH64910-AAH85051) and oligonucleotide sequences  
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid  
 CC molecules and proteins of the invention are useful for diagnosis and  
 CC monitoring viral infection, in screens for antiviral agents and for  
 CC monitoring viral gene expression or activity during a treatment regimen.  
 CC The nucleic acid molecules are also useful as antisense constructs to  
 CC control viral gene expression in infected cells and tissues and to create  
 CC transgenic viral resistant shrimp.  
 XX Sequence 922 AA;  
 SQ  
 Query Match 11.5%; Score 86; DB 22; Length 922;  
 Best Local Similarity 21.9%; Pred. No. 10;  
 Matches 39; Conservative 28; Mismatches 63; Indels 48; Gaps 7;  
 QY 3 VVTWIONGKMSSTIVSEEDFLPVYKGELEKGYQDFDWEISGP--EGKDKAGYVNLK 34  
 DB 310 VTESMRNGCCCFKNDKWLAKRESNLNTVFGEDDESKAYAVSDSEDEDENEEEV 369  
 QY 35 GYQDFGWEISGEGKDGAGVNLKSDTFIKPV-FKKIEKKKEENKPTFDVSKKKNPQ 93  
 DB 370 DDYNNETIESSVG-----NVIKNLIRKIGLSDVEEKEEGEQSEEEEDSDDDDD 421  
 QY 94 VNHSQNLNESHKEDL-----QREHSQKSDSTKQVATVL---DKNNISSKSTNN 141  
 DB 422 DASSVCSSSSSSVTVAAAEDEEDEDKDKD-KATVWEDEDDKESVISSSSDS 478  
 RESULT 36  
 AAY08642  
 ID AAY08642 standard; Protein; 1315 AA.  
 AC AAY08642;  
 XX 20-MAR-2003 (updated)  
 DT 09-AUG-1999 (first entry)  
 XX S. aureus SdrD protein.  
 DE  
 XX Fibrinogen-binding protein; alpha chain; beta chain; ClfB; SdrC; SdrD;  
 KW SdrE; fibrinogen; medical device; competitive inhibitor; pharmaceutical;  
 KW treatment; infection; septicemia; osteomyelitis; mastitis; endocarditis;  
 KW extracellular matrix; vascular graft; vascular stent; vaccine;  
 KW intravenous catheter; artificial heart valve; cardiac assist device;  
 KW antibacterial.  
 XX Staphylococcus aureus.  
 OS  
 XX WO9927109-A2.  
 PN  
 XX 03-JUN-1999.  
 PD  
 XX 25-NOV-1998; 98WO-0525246.  
 XX 31-AUG-1998; 98US-0098427.  
 PR 26-NOV-1997; 97US-0066815.  
 XX (INHIT-) INHIBITEX INC.  
 PA

PA (FORF-) FORFAS T/A BIORESEARCH IRELAND.  
 PA (TEXA) UNIV TEXAS A & M.  
 PA (PATT/) PATTI J M.  
 PA (POST/) FOSTER T J.  
 PA (JOSE/) JOSEFSSON E.  
 PA (EIDH/) EIDHIN D N.  
 PA (HOOK/) HOOK M A O.  
 PA (PERK/) PERKINS S E.  
 XX Eidhin DN, Foster TJ, Hook MAO, Joseffsson E, Patti JM;  
 PI Perkins SE;  
 XX WPI: 1999-357844/30.  
 DR N-PSDB; AAX77593.  
 XX Staphylococcus aureus fibrinogen-binding proteins for treating  
 PT septicemia, osteomyelitis, mastitis or endocarditis  
 XX Claim 8; Fig 8; 143pp; English.  
 PS This invention describes novel Staphylococcus aureus fibrinogen-binding  
 CC proteins that bind both the alpha and beta fibrinogen chains. The  
 CC proteins (and their encoding nucleic acids are ClfB, SdrC, SdrD and  
 CC SdrE). Staphylococcus aureus is thought to utilize fibrinogen to adhere  
 CC to medical devices, binding proteins that bind both the alpha and beta  
 CC fibrinogen chains (ClfB, SdrC, SdrD and SdrE) can therefore be used as  
 CC competitive inhibitors to block this binding. Antibodies against ClfB,  
 CC SdrC, SdrD and SdrE inhibit ClfB, SdrC, SdrD and SdrE mediated binding.  
 CC The proteins of the invention can be used in a pharmaceutical composition  
 CC for the treatment of Staphylococcus aureus infection e.g. septicemia,  
 CC osteomyelitis, mastitis or endocarditis or to inhibit the binding of  
 CC S. aureus to the extracellular matrix. The proteins or their fragments  
 CC may be used to coat a medical device to reduce the S. aureus infection of  
 CC an indwelling medical device, especially where the medical device is  
 CC selected from the group consisting of vascular grafts, vascular stents,  
 CC intravenous catheters, artificial heart valves, and cardiac assist  
 CC devices. ClfB, SdrC, SdrD, SdrE, or an active fragment, subdomain or  
 CC encoding gene may be used as a vaccine. The DS (aspartate serine) repeat  
 CC region or a gene encoding it may be used as an identifying probe for the  
 CC identification of genes and encoding proteins from Staphylococcus aureus  
 CC (other than ClfA), S. hemolyticus, S. lugdenensis, and S. schleriferi.  
 CC The proteins of the invention have antibacterial activity.  
 CC (Updated on 20-MAR-2003 to correct PA field.)  
 XX Sequence 1315 AA;  
 SQ  
 Query Match 11.5%; Score 86; DB 20; Length 1315;  
 Best Local Similarity 24.2%; Pred. No. 17;  
 Matches 36; Conservative 22; Mismatches 65; Indels 26; Gaps 4;  
 QY 3 VVTWIONGKMSSTIVSEEDFLPVYKGELEKGYQDFDWEISGP--EGKDKAGYVNLK 60  
 DB 819 VVTILKN-----ENGEVLQTTKDKGKYQFTGLENGTYKVEFETPSGYTPT--- 865  
 QY 61 DTFIKPVFKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 866 -----QVSGTGDEGIDSGTSTGVKDKNDTDSGTGFKYPTNLGDYVWEDTKNGVQD 920  
 QY 121 KD-----VTATVLDKNNISSKSTNNPN 143  
 DB 921 KDEKGISGVTVTLKDENDKVLKTVTTDEN 949  
 RESULT 37  
 ABJ18969  
 ID ABJ18969 standard; Protein; 1315 AA.  
 XX ABJ18969;  
 AC ABJ18969;  
 XX 06-MAR-2003 (first entry)  
 DT Pathogen specific antigen related staphylococcal protein SEQ ID No 124.  
 XX





QY 121 KD-----VTATVLDKNNISSKSTNNPN 143  
 DB 921 KDEKIGSGVTTLKDNENDKVLKVTITDEN 949

RESULT 39  
 AAU37544  
 ID AAU37544 standard; Protein; 1349 AA.  
 AC AAU37544;  
 DT 14-FEB-2002 (first entry)  
 DE Staphylococcus aureus cellular proliferation protein #1714.  
 KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 OS Staphylococcus aureus.  
 XX WO200170955-A2.  
 XX 27-SEP-2001.  
 XX 21-MAR-2001; 2001WO-US09180.  
 XX 21-MAR-2000; 2000US-191078P.  
 XX 23-MAY-2000; 2000US-206848P.  
 XX 26-MAY-2000; 2000US-207727P.  
 XX 23-OCT-2000; 2000US-242578P.  
 XX 27-NOV-2000; 2000US-253625P.  
 XX 12-DEC-2000; 2000US-257931P.  
 XX 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 XX N-PSDB; AAS55403.  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 XX Example 3; Seq ID No 13137; 51pp; English.  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 1349 AA;  
 Query Match 11.5%; Score 86; DB 22; Length 1349;  
 Best Local Similarity 24.2%; Pred. No. 17;  
 Matches 36; Conservative 22; Mismatches 65; Indels 26; Gaps 4;

QY 3 VTTVTIONGKEMSSSTIVSEEDFILPVYKGELEKGYQDCWEISGF--EGKXDAGYVINLSK 60  
 DB 819 VTTVLKN-----ENGEVLQTTKTDKQKGYQFTGLENGTYKVEFETSGYTPPT--- 865

QY 61 DTFIKPVPFKKIEEKEENKPTFDVSKKKDPQVNHSQLNESHKREDLQREHESQKSDST 120  
 DB 866 -----QVSGTDEGIDSNGTSTTGVIKDKNDTIDSGFYKPTYNLGDYVWEDTNKNGVQD 920

QY 121 KD-----VTATVLDKNNISSKSTNNPN 143  
 DB 921 KDEKIGSGVTTLKDNENDKVLKVTITDEN 949

RESULT 40  
 AAB23563  
 ID AAB23563 standard; Protein; 1980 AA.  
 AC AAB23563;  
 XX 09-JAN-2001 (first entry)  
 DE Human sodium channel 8A (SCN8A) protein sequence.  
 XX Sodium channel 8A; SCN8A; human; excitatory cell.  
 XX Homo sapiens.  
 XX JP2000201684-A.  
 XX 25-JUL-2000.  
 XX 11-JAN-1999; 99JP-0004645.  
 XX 11-JAN-1999; 99JP-0004645.  
 XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
 XX WPI; 2000-545978/50.  
 XX N-PSDB; AAA93791.  
 XX New sodium channel SCN8A useful for the elucidation physiological  
 PT mechanism participated by excitatory cells -  
 XX  
 XX Disclosure; Page 5-12; 14pp; Japanese.  
 CC This sequence represents the human sodium channel SCN8A protein. The  
 CC invention also provides the SCN8A encoding gene sequence and an antibody  
 CC against the sodium channel. The SCN8A protein alpha subunit and its  
 CC encoding cDNA can be used in the elucidation of the physiological  
 CC mechanisms of excitatory cells.  
 XX  
 XX Sequence 1980 AA;  
 Query Match 11.5%; Score 86; DB 21; Length 1980;  
 Best Local Similarity 26.2%; Pred. No. 29;  
 Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;

QY 37 QPDGWEISGFEGKDGAGYVINLSKDTFIKVPFKKIEEKEENKPTFDVSKKKDPQVNH 96  
 DB 993 ENNNLQISVIRIKKGAVNT-KLVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1048

QY 97 SOLNESHKREDLQREHESQKSDSTKDVATVLDKNNISSKSTNNPN 143  
 DB 1049 TGA-DIHRNGDFQKNGNGTTSIGSGSVEKYIIDEDHM---SFINNPN 1091

Search completed: February 10, 2004, 10:53:53  
 Job time : 33.9025 secs



```
Db 1150 SHDKRPFPSKVE-----QKSSSRKSDNDKDLTHLDFVQNNFSEIFMKNLLSP 1201
QY 143 NK 144
Db 1202 QK 1203

RESULT 2
CYL2_HUMAN
ID -CYL2_HUMAN STANDARD; PRT; 348 AA.
AC Q1403;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cyclicin II (Multiple-band polypeptide II).
GN CYLC2 OR CYL2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95255491; PubMed=7737358;
RA Hess H., Heid H., Zimbelmann R., Franke W.W.;
RT "The protein complexity of the cytoskeleton of bovine and human sperm
RT heads: the identification and characterization of cyclicin II.";
RL Exp. Cell Res. 218:174-182(1995).
CC -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -!- TISSUE SPECIFICITY: Testis.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46788; CAA86752.1; -.
DR PIR; I37271; I37271.
DR Genew; HGNC:2583; CYLC2.
DR MIM; 604035; -.
DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 25 347 31 X 3 AA REPEATS OF K-K-X.
FT REPEAT 157 240 3 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 157 184 1.
FT REPEAT 185 212 2.
FT REPEAT 213 240 3.
SQ SEQUENCE 348 AA; 39079 MW; D86765599C1809E7 CRC64;

Query Match 12.7%; Score 95; DB 1; Length 348;
Best Local Similarity 30.9%; Pred. No. 2.1;
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 29 KGELEKGYQFDGWEISGFEGKKDAGVNLKDTFTKPVFKIEKKEENKPTP---DV 85
Db 205 ESEGEKG---GTEKDSKKGKDS-----KKGKDSALQLQAVKDEKDDGKKDKGDE 256

QY 86 SK--KKDNPQVNHSQLN-----ESHKREDLQREHSHOKSDSTKD---VTATVLDKNI 133
Db 257 SKDAKKDAKEIKGKKDKKKPSTSDSDSKDDVKE---SKDKATKDAKKVAKDTKEKSA 313

QY 134 SSK 136
Db 314 DSK 316

RESULT 3
```

```
TOP2_PLAFK
ID TOP2_PLAFK STANDARD; PRT; 1398 AA.
AC P41001;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94316496; PubMed=8041616;
RA Chessman S., McAleese S., Goman M., Johnson D., Horrocks P.,
RA Ridley R.G., Kilbey B.J.;
RT "The gene encoding topoisomerase II from Plasmodium falciparum.";
RL Nucleic Acids Res. 22:2547-2551(1994).
CC -!- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X79345; -. NOT_ANNOTATED_CDS.
DR HSSP; P06786; 1BGW.
DR InterPro; IPR003594; ATPbind ATPase.
DR InterPro; IPR003957; CBFA_NFVB_topis.
DR InterPro; IPR001241; DNA_topoisolv.
DR InterPro; IPR002205; DNA_topoisolv.
DR Pfam; PF00204; DNA_gyraseB; 1.
DR Pfam; PF00521; DNA_topoisolv; 1.
DR Pfam; PF02518; HATPase_c; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PRINTS; PR00418; TP12FAMILY.
DR ProDom; PD000742; DNA_topoisolv; 1.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00433; TOP2c; 1.
DR SMART; SM00434; TOP4c; 1.
DR PROSITE; PS00177; TOPOISOMERASE II; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
FT NP_BIND 144 149 ATP (POTENTIAL).
FT ACT_SITE 830 830 DNA CLEAVAGE (BY SIMILARITY).
FT DOMAIN 271 281 POLY-ASN.
FT DOMAIN 308 316 POLY-ASN.
FT DOMAIN 1089 1093 POLY-LYS.
FT DOMAIN 1227 1234 POLY-LYS.
SQ SEQUENCE 1398 AA; 161029 MW; BAAD7BEE80FE5BE9 CRC64;

Query Match 12.7%; Score 95; DB 1; Length 1398;
Best Local Similarity 23.8%; Pred. No. 9.4;
Matches 36; Conservative 36; Mismatches 53; Indels 26; Gaps 6;

QY 9 NGKEMSTTVSEEDFIL--PVYKGELEKGYQFDGWEISGFEGKKDAGVNLKDTFTK 66
Db 1133 NEEIAGITVKDYDYLKSMPIFSLTLEK---VEDLLTQLKEKELEILRNITVTWMLK 1189

QY 67 VFKKIEE-----KKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSHQ 115
```

Db 1190 DIEKVEAIEPQNVNELNRESNK--PKVARKQ-----GPSMKKKKKKKLSSDEESE 1242

Oy 116 ---KSDSTKVATVLDKNNISSKSTNNPN 143

Db 1243 GGDTSDSSEFLVTLNLIKNTKTTSSNN 1273

RESULT 4

MAPB HUMAN

ID MAPB HUMAN STANDARD; PRT; 2468 AA.

AC P46821;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain LC1].

DE LC1].

GN MAP1B.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

EN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RX MEDLINE=95104835; PubMed=7806212;

RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;

RT "Cloning of human microtubule-associated protein 1B and the identification of a related gene on chromosome 15.";

RT Genomics 22:273-280(1994).

CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.

CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN STABILIZING MICROTUBULES.

CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE WITH MAP1A AND MAP1B PROTEINS.

CC -!- DOMAIN: Has a highly basic region with many copies of the sequence KKEE and KKEI/V repeated but not at fixed intervals, which is responsible for the binding of MAP1B to microtubules.

CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated from MAP1B by proteolytic processing. It is free to associate with both MAP1A and MAP1B. It interacts with the amino-terminal region of MAP1B (By similarity).

CC -!- SIMILARITY: TO MAP1A.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; L06237; AAA18904.1; .

DR Genew; HGNC:6836; MAP1B.

DR MIM; 157129; .

DR GO; GO:0005875; C:Microtubule associated complex; TAS.

DR InterPro; IPR000102; MAP1B neuraxin.

DR Pfam; PF00414; MAP1B neuraxin; 10.

DR PROSITE; PS00230; MAP1B NEURAXIN; 6.

KW Microtubules; Repeat; Phosphorylation.

FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.

FT REPEAT 1878 1894 MAP1B 1.

FT REPEAT 1895 1911 MAP1B 2.

FT REPEAT 1912 1928 MAP1B 3.

FT REPEAT 1929 1945 MAP1B 4.

FT REPEAT 1946 1962 MAP1B 5.

FT REPEAT 1963 1979 MAP1B 6.

FT REPEAT 1997 2013 MAP1B 7.

FT REPEAT 2014 2030 MAP1B 8.

FT REPEAT 2031 2047 MAP1B 9.

FT REPEAT 2048 2064 MAP1B 10.

FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY KKEE AND KKEI/V REPEATS).

FT SQ SEQUENCE 2468 AA; 270618 MW; 540839CDBFD9D461 CRC64;

Query Match 12.4%; Score 92.5; DB 1; Length 2468;

Best Local Similarity 31.0%; Pred. No. 26;

Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Oy 58 LSKDTFIKVPFKIEEKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQRE-----E 112

Db 638 VKKETVKP-----EDKKEEKEKPKKEVAKEDKTPi---KKEEKPKKEVKKEIK 689

Oy 113 HSQKSDSTKDV 123

Db 690 KEEKKPKKEV 700

RESULT 5

SPT7 YEAST

ID SPT7 YEAST STANDARD; PRT; 1332 AA.

AC P35177;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Transcriptional activator SPT7.

GN SPT7 OR YBR081C OR YBR0739.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

EN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=S288C;

RX MEDLINE=95225044; PubMed=7713415;

RA Gansharoff L.J., Dollard C., Tan P., Winston F.;

RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein important for transcription in vivo.";

RL Genetics 139:523-536(1995).

EN [2]

RP SEQUENCE FROM N.A.

RX STRAIN=S288C;

RX MEDLINE=95076715; PubMed=7985423;

RA van der Aart Q.J.M., Barthe C., Doignon F., Aigle M., Crouzet M., Steensma H.Y.;

RT "Sequence analysis of a 31 kb DNA fragment from the right arm of Saccharomyces cerevisiae chromosome II.";

RL Yeast 10:959-964(1994).

EN [3]

RP SEQUENCE OF 1-835 FROM N.A.

RX STRAIN=S288C;

RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestazu A., Vissers S.;

RT Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.

EN [4]

RP SEQUENCE OF 463-523 FROM N.A.

RX MEDLINE=92285152; PubMed=1350857;

RA Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J., David I.B.;

RT "The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins.";

RL Nucleic Acids Res. 20:2603-2603(1992).

CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY OTHER GENES.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- SIMILARITY: Contains 1 bromodomain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; L22537; AAC37424.1; -
DR EMBL; X76294; CAA53940.1; -
DR EMBL; Z35950; CAA85026.1; -
DR EMBL; M87651; AAA35087.1; -
DR PIR; S41552; S41552.
DR HSSP; Q2831; 1B91.
DR SGD; S0000285; SPT7.
DR GO; GO:000124; C:SAGA complex; IDA.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS00633; BROMODOMAIN 1; 1.
DR PROSITE; PS00114; BROMODOMAIN 2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain.
FT DOMAIN 458 528 BROMODOMAIN
FT SEQUENCE 1332 AA; 152616 MW; 083B6324669244F CRC64;

Query Match 12.1%; Score 90.5; DB 1; Length 1332;
Best Local Similarity 24.5%; Pred. No. 19;
Matches 39; Conservative 26; Mismatches 57; Indels 37; Gaps 8;

QY 5 VTIONGKMSSTIVSEEDPILPVYKGELEKGYQFDGWEISGEGKKDAGYVNLKSDTFI 64
Db :||: :| :|| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
549 IITRNADLEKEI---EDM-----EKDKDYELDEEEVAGSRKG----LNMGAHMLA 594

QY 65 K---PVFKTEEEKKEENKPTD-----VSKKKD-----NPDVNHLSQLNESH 103
Db :||: :| :|| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
595 KENGKVKSEKSKTKVDAPTNDKLTSVIPEGEKEKKTASTVTIVHENVKNEIKENG 654

QY 104 RKEDLQR-BEHSQKSDSTKDTVTATVLD-KNNISSKSTTN 140
Db :||: :| :|| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
655 KNEEQDMVESSKTEDESSKDAADAAKKDTEGLQDKTAEN 693

RESULT 6
YDT2 SCHPO
ID YDT2 SCHPO STANDARD; PRT; 1888 AA.
AC Q14207;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C6B12.02c in chromosome I.
GN SPAC6B12.02c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Gouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Ruckie E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Mottier S.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

```

```

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Usery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z98531; CAB11064.1; -
DR PIR; T39009; T39009.
DR GenBank; SPombe; SPAC6B12.02c; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 697 717 POTENTIAL.
FT TRANSMEM 755 775 POTENTIAL.
FT TRANSMEM 866 886 POTENTIAL.
FT TRANSMEM 915 935 POTENTIAL.
FT TRANSMEM 977 997 POTENTIAL.
FT TRANSMEM 1033 1053 POTENTIAL.
FT TRANSMEM 1336 1356 POTENTIAL.
FT TRANSMEM 1645 1665 POTENTIAL.
FT DOMAIN 1662 1665 POLY-LEU.
FT SEQUENCE 1888 AA; 217432 MW; 8AD3BBCE32397C29 CRC64;

Query Match 11.9%; Score 89; DB 1; Length 1888;
Best Local Similarity 23.9%; Pred. No. 36;
Matches 37; Conservative 29; Mismatches 59; Indels 30; Gaps 7;

QY 14 SSTIVSEEDF-----ILPVY---KGELEKGYQFDGWEISGFE-----GKKDAG 53
Db :||: :| :|| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
390 SSSLTSENPQLNVAANAVSTIPVYRTTKTKMKKN-RFKYVEVEKLPDLILESYGKKAPK 448

QY 54 YVINLSKDTFIKPVFKTEEEKKEENKPTFDVSKKKDNPQVNHLSQLNESHREDLQREH 113
Db :||: :| :|| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
449 FURVFARSSSHIP--KMTRRKQMDSKYFSPDKESDQVLDVLSDWYSGKHVLVQSH 506

QY 114 SOKSDS-TKDVATVLDKN-----NISKSTTN 141
Db :||: :| :|| :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
507 SYKKPSDSKSGVGNIFSVNKKHGSVNNAKTAANN 541

RESULT 7
IGA2 HAEIN
ID IGA2 HAEIN STANDARD; PRT; 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / Serotype B;
RX MEDLINE=92234949; PubMed=1373717;
RA Paulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.

```

```
[2]
SEQUENCE OF 1-94 FROM N.A., AND MUTAGENESIS OF SER-84.
MEDLINE=92102204; PubMed=1662027;
RA Sreedharan S., Peterson L.R., Fisher L.M.;
RT "Ciprofloxacin resistance in coagulase-positive and -negative
RT staphylococci: role of mutations at serine 84 in the DNA gyrase A
RT protein of Staphylococcus aureus and Staphylococcus epidermidis.";
RL Antimicrob. Agents Chemother. 35:2151-2154(1991).
CC -!- FUNCTION: DNA GYASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
CC ENZYME FORMS AN A2B2 TETRAMER.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
-----
CC EMBL; AB016744; AAC03602.1; -.
DR EMBL; S72603; AAB20672.1; -.
DR PIR; A49832; A49832.
DR HSSP; P09097; LABA.
DR InterPro; IPR002205; DNA_topoisolv.
DR Pfam; PF03989; DNA_gyraseA_C; 6.
DR ProDom; PD000742; DNA_topoisoIV; 1.
DR SMART; SM00434; TOP4c; 1.
DR TIGREMS; tIGR01063; gyrA; 1.
DR Topoisomerase; Isomerase; DNA-binding; Antibiotic resistance;
KW Complete proteome.
FT ACT_SITE 123 123 DNA CLEAVAGE (BY SIMILARITY).
FT MUTAGEN 84 84 S->F: RESISTANT TO CIPROFLOXACIN.
SQ SEQUENCE 893 AA; 100113 MW; 2A6A7CD345A526CE CRC64;

Query Match 11.6%; Score 87; DB 1; Length 893;
Best Local Similarity 25.3%; Pred.No.23;
Matches 38; Conservative 27; Mismatches 65; Indels 20; Gaps 7;

QY 4 TVTIQNGKMS-STIVSEEDFILPVYKGLEKGYQFDGWEISGFEGKKDAGY-VINLSKD 61
Db ||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:|||||
746 TITERNGNVCITTVTGEEEDLMVTVNAVGI---IRLDVHDIQ-NGRAAQGVFLMKLGDG 801

QY 62 FTIKPFVKKIEEKKEEN-----KPTDFYSKKKNPQVNHSQLNSHRKEDIQREE 112
Db ||:||:||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
802 QFVSITVAKVNEEDNEENADEAQQSSTTTETADVEEVVD---DOTPGNAIHTEGDAMES 857

QY 113 -HSQKSDSKTDVTATVLDKNNISSKSTTN 141
Db ||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:|||||
858 VESPENDDRIDIRQDPMDRVNEDIESASDN 887

RESULT 9
MAP_MOUSE STANDARD; PRT; 2464 AA.
ID MAP_MOUSE
AC P14873;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
DE [Contains: MAP1 light chain LC1].
DE MAP1B OR MTAP1B OR MTAP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Eutelestomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```





KW Nuclear protein; Hydrolase; Nuclease; Exonuclease; Repeat.  
 FT DOMAIN 525 528 POLY-GLU.  
 FT DOMAIN 793 797 POLY-ASN.  
 FT DOMAIN 955 959 YSGN REPEATS.  
 FT MUTAGEN 693 693 Y->H: IN ALLELE TP1-1; ACTIVATES  
 FT TRANSCRIPTION OF THE PROMOTER-DEFECTIVE  
 FT YEAST SUP4 TRNA(TYR) ALLELE SUP4A53T61.  
 SQ SEQUENCE 1006 AA; 115933 MW; 5DD5B0245F3E12A CRC64;  
 Query Match 11.6%; Score 86.5; DB 1; Length 1006;  
 Best Local Similarity 19.0%; Pred. No. 29;  
 Matches 32; Conservative 40; Mismatches 47; Indels 49; Gaps 8;  
 QY 7 IQNGKEMSTIVSEDFILPVY--KGELEKGYQFDGWEISGPEGKDGAGYVNLKDTFI 64  
 DB 415 MSKGDRHPTVATEQ--LQMYDTQGNLAG--SWNLITSD-----WVRLKKEML 460  
 QY 65 -----KPVFKKIEEKEENKPTFDVSKKONPQVNHQ-----LNESHK 105  
 DB 461 ANEGNEEAIAKVQKQSDKNELMKDISKEEDDAVSKANKTNFNLAEVNMKQIKIINKHRL 520  
 QY 106 EDLQREEH-----SQKSDSTKDTATVLDK-----NNISSKSTT 139  
 DB 521 EKONEEBEIAKSKVKTEKAESCDLDAETKDEIVADVNDRENSETT 568  
 RESULT 11  
 ID GLNA\_PYRKO STANDARD; PRT; 443 AA.  
 AC 008467;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
 GN GLNA.  
 OS Pyrococcus kodakaraensis.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Thermococcus.  
 OX NCBI\_TaxID=69014;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KOD1;  
 RX MEDLINE=97316461; PubMed=9172372;  
 RA Adul Rahman R.N.Z., Jongsareejit B., Fujiwara S., Imanaka T.;  
 RT "Characterization of recombinant glutamine synthetase from the  
 RT hyperthermophilic archaeon Pyrococcus sp. strain KOD1.";  
 RL Appl. Environ. Microbiol. 63:2472-2476 (1997).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
 CC L-glutamine.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D86222; BAA20530.1; -.  
 DR HSP; P06201; ILGR.  
 DR InterPro; IPR001691; GLN synth.  
 DR InterPro; IPR004809; GLNA.  
 DR InterPro; IPR001637; GLNA\_adenyltn.  
 DR Pfam; PF00120; gln-synt\_1.  
 DR Pfam; PF03951; gln-synt\_N; 1.  
 DR ProDom; PD001057; Gln synt C; 1.  
 DR TIGRFAMs; TIGR00653; GLNA; 1.  
 DR PROSITE; PS00180; GLNA\_1; 1.  
 DR PROSITE; PS00181; GLNA\_ATP; 1.  
 KW Ligase.

FT BINDING 362 362 AMP (UNDER CONDITIONS OF ABUNDANT  
 FT POLY-GLU.  
 SQ SEQUENCE 443 AA; 50259 MW; 9426DCCPEEF18168 CRC64;  
 Query Match 11.5%; Score 86; DB 1; Length 443;  
 Best Local Similarity 17.8%; Pred. No. 13;  
 Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;  
 QY 28 YKGELEKGYQFDGWEISGPEGKDGAGYVNLKDTFI----- 64  
 DB 42 YEEAEDGVSGFDGSGSIPGEGIEDSLFKADPSTYAEIPWEGIGRVYGYIKGDEPYQA 101  
 QY 65 --KPVFKKIEEKEEN-----KPTFDVSKKXD----- 90  
 DB 102 DPGILKRVLRLEKEGLKAHIGPEPEFYIEKKGNWELHIPDSGGYFDLVGLDKAREIR 161  
 QY 91 -----NQVNHSQLNESHKEDLQREHSQKSD---STKDTATVLD 129  
 DB 162 REIALYMPYLGKPEVLHHEVGKAQHEIDFDYDEALRTADNIVSPKVVKAAVE 215  
 RESULT 12  
 TIG\_STAM  
 ID TIG\_STAM STANDARD; PRT; 433 AA.  
 AC Q99TIG;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trigger factor (TF).  
 GN TIG OR SAVI675 OR SAI499 OR MW1619.  
 OS Staphylococcus aureus (strain N315) / ATCC 700699).  
 OS Staphylococcus aureus (strain N315), and  
 OS Staphylococcus aureus (strain MW2).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=158878, 158879, 196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MU50 / ATCC 700699, and N315;  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Negai Y., Lian J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hotoyama A.,  
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 RT aureus.";  
 RL Lancet 357:1225-1240 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MW2;  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiramatsu K.;  
 RT "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA.";  
 RL Lancet 359:1819-1827 (2002).  
 CC -!- FUNCTION: Involved in protein export. Acts as a chaperone by  
 CC maintaining the newly synthesized protein in an open conformation  
 CC (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PI3ASE FAMILY. TIG SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AP003363; BAB57837.1; -.

```
DR EMBL; AP003134; BAB42766.1; -.
DR EMBL; AP004827; BAB95484.1; -.
DR PIR; A89951; A89951.
DR HAMAP; MF 00303; -.
DR InterPro; IPR001179; FKBP_PP1ase.
DR InterPro; IPR005215; Trig_fac.
DR Pfam; PF00254; FKBP; 1.
DR TIGRFAMs; TIGR00115; tlg; 1.
DR PROSITE; PS00453; FKBP_PP1ASE_1; FALSE NEG.
DR PROSITE; PS00454; FKBP_PP1ASE_2; FALSE NEG.
DR PROSITE; PS00059; FKBP_PP1ASE_3; 1.
DR Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.
KW DOMAIN 163 248 PP1ase, FKBP-TYPE.
FT SEQUENCE 433 AA; 48609 MW; 8865D9AF6A1BC1E7 CRC64;

Query Match 11.4%; Score 85.5; DB 1; Length 433;
Best Local Similarity 22.9%; Pred. No. 14;
Matches 49; Conservative 30; Mismatches 52; Indels 83; Gaps 13;

QY 3 VTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 34
DB 93 VSVTQIEKGKDFTEATVPEVEVGLDYKGLGIEKQETELSDDELQEAIDHSLGLAEM 152
QY 35 -----GYQFGWELSG--FEGKKDAGYVNLKDTFKVPK-----KIE 72
DB 153 VKKEDGVNGDTVNI-DFSG-SVDGEFEGGQAGYDLGIGSGFT-PGFEEQLGKMYD 210
QY 73 EKKE-----EE--NKPTFVDS-----KKDNQVNHSQLNE-----SHRKE 106
DB 211 EKDQVVVTPPEVHAEELAGKATFKYKNEIKFEVPELTDEIANELDAEANTVDVEYKE 270
QY 107 DLQREHSQKSDTKDVTATVLDKNNISSKSTTN 140
DB 271 NLRKRLAEQKATDAENV-----EKEAITKATDN 299

RESULT 13
DPOM ASCIM STANDARD; PRT; 1202 AA.
AC P22374;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable DNA polymerase (EC 2.7.7.7).
OS Ascobolus immersus.
OG Mitochondrion.
OC Plasmid pA12.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;
OC Pezizales; Ascombolaceae; Ascobolus.
OX NCBI_TaxID=5191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2/I;
RX MEDLINE=90066356; PubMed=2573821;
RA Kempken F., Meinhardt F., Esser K.;
RT "In organello replication and viral affinity of linear,
RT extrachromosomal DNA of the ascomycete Ascobolus immersus.";
RL Mol. Gen. Genet. 218:523-530(1989).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -!- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG, TO
CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; X15982; CAA34106.1; -.
DR PIR; S05362; S05362.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR004868; DNA_pol_B_2.
DR Pfam; PF03175; DNA_pol_B_2; 1.
DR SMART; SM00486; POLBc_1.
DR PROSITE; PS00116; DNA-POLYMERASE_B; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
KW DNA-binding; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Plasmid; Mitochondrion.
FT SEQUENCE 1202 AA; 138279 MW; 51D41FCEBDBF2CDE CRC64;

Query Match 11.4%; Score 85.5; DB 1; Length 1202;
Best Local Similarity 22.5%; Pred. No. 41;
Matches 36; Conservative 31; Mismatches 64; Indels 29; Gaps 6;

QY 4 TVTIQNGKEMSGSTIVSEEDFILPVYKGBLEK-----GYQFGWELSGFEGKKDAGYVNL 57
DB 344 TLAIFLEREDITHVMSYDE-----GDIDSKFPKGSLSFDFKPLKTIETGKYANYTFP 396
QY 58 LSKOTFIKVPFKIE-----EKKEENK-PTFVDSKKDNQVNHSQLNESHKEDLQR 110
DB 397 IKKDIWVKDINKINFNGLDLPKTMDSLKPWNLKLNKDKTSGEIRMTIKNNKNSQSYDI-- 454
QY 111 EHSQKSDTKDVTATVLDKNNI-----SKSSTTNPN 143
DB 455 IGHMIINDGENVITPNRAVDNSIIKIFTVTDMSGNTNDPN 494

RESULT 14
IF2_STAEP STANDARD; PRT; 720 AA.
ID IF2_STAEP
AC Q8CST4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-2.
OS INF8 OR SE0945.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (by similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE016747; AA004542.1; -.
CC HAMAP; MF 00100; -.
CC Pfam; PF00009; GTP_EFTU_1.
CC Pfam; PF03144; GTP_EFTU_D2; 1.
CC Pfam; PF04760; IF2_N_2.
CC Pfam; PF04760; IF2_N_2.
CC ProDom; PD186100; IF2; 1.
CC TIGRFAMs; TIGR00487; IF-2; 1.
CC TIGRFAMs; TIGR00231; small_GTP; 1.
CC PROSITE; PS01176; IF2; 1.
CC Initiation factor; Protein biosynthesis; GTP-binding;
CC Complete proteome.
KW
```

```
FT DOMAIN 225 373 G-DOMAIN
FT NP_BIND 231 238 GTP (BY SIMILARITY)
FT NP_BIND 277 281 GTP (BY SIMILARITY)
FT NP_BIND 331 334 GTP (BY SIMILARITY)
SQ SEQUENCE 720 AA; 79343 MW; 07FBSA6A59CF970C CRC64;

Query Match 11.4%; Score 85; DB 1; Length 720;
Best Local Similarity 28.0%; Pred. No. 26;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

Qy 71 IEEKKEENKTFDVSKKD-----NPQVNHSHKEDLQREHSHQKSDSTK-----D 122
Db 38 LEEQIKALDKKFKASQAKDKNTQNNHQSNNKQNSNDKQKQNSKPTKKKEQN 97

Qy 123 VTATVLDKNNISSKSTNNPNK 144
Db 98 NKGQKQNNKNTKNNKNNK 119

RESULT 15
ID _IGA0_HAEN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA OR IGAL OR HI0990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype D;
RA Wright A., Fishman Y., Tai F., Plaut A.G.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; X59800; -, NOT ANNOTATED CDS.
DR EMBL; U32779; AAC22651.1; -.
DR PIR; H64106; H64106.
DR MEROPS; S06.001; -.
DR TIGR; HI0990; -.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertactin.
DR Pfam; PF03797; Autotransporter; 1.
DR Pfam; PF02395; IGAL; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR TIGRFAMs; TIGR01414; autotrans barl; 1.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal;
KW Complete proteome. 25
FT SIGNAL 1
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT CONFLICT 253 254 EN -> GV (IN REF. 1).
FT CONFLICT 272 272 G -> A (IN REF. 1).
FT CONFLICT 464 464 G -> E (IN REF. 1).
FT CONFLICT 866 866 S -> T (IN REF. 1).
FT CONFLICT 1036 1036 A -> D (IN REF. 1).
FT CONFLICT 1074 1074 A -> G (IN REF. 1).
FT CONFLICT 1421 1421 A -> G (IN REF. 1).
FT CONFLICT 1545 1545 H -> T (IN REF. 1).
SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;

Query Match 11.3%; Score 84.5; DB 1; Length 1694;
Best Local Similarity 26.1%; Pred. No. 70;
Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

Qy 56 INLSKDTFIKPVFKIEEKKEENKPTFDVSKKDNQVNHSHKEDLQREHSH- 114
Db 1288 INTGSATAITETAESKDKPTETAASTEDASQHKANTVADNSVANNSESDPKSRRRSI 1347

Qy 115 --QKSDSTKDVATVLDKNNISSKSTNNPNK 144
Db 1348 SQPQETSAEETTAASTDETTIADNSKRSKPNR 1379

RESULT 16
YFK8_YEAST
ID YFK8_YEAST STANDARD; PRT; 778 AA.
AC P43610;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 88.7 kDa helicase in CDC26-SAP155 intergenic region.
GN YFK038W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=96287654; PubMed=8686381;
RA Eki T., Naitou M., Hagiwara H., Abe M., Ozawa M., Sasanuma S.-I.,
RA Sasanuma M., Tsuchiya Y., Shibata T., Watanabe K., Ono A.,
```

```

RA Yamazaki M.-A., Tashiro H., Hanaoka F., Murakami Y.;
RT "Fifteen open reading frames in a 30.8 kb region of the right arm of
RL chromosome VI from Saccharomyces cerevisiae.";
RC Yeast 12:177-190(1996).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D50617; BAA09277.1; -.
CC PIR; S56293; S56293.
CC SGI; S0001934; YFR038W.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR000330; SNF2_N.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF00176; SNF2_N; 1.
CC SMART; SM00487; DEXDC; 1.
CC SMART; SM00490; HELICG; 1.
KW Hypothetical protein; Nuclear protein; DNA-binding; Helicase;
KW ATP-binding.
FT NP_BIND 247 254 ATP (POTENTIAL).
FT SITE 352 355 DSGH BOX.
FT SITE 778 AA; 88730 MW; 36C0857B5EABD84 CRC64;
SQ SEQUENCE 778 AA; 88730 MW; 36C0857B5EABD84 CRC64;

Query Match 11.2%; Score 84; DB 1; Length 778;
Best Local Similarity 25.5%; Pred. No. 33;
Matches 37; Conservative 18; Mismatches 56; Indels 34; Gaps 5;

QY 13 MSSTIVSEDFILPVYKGEKGYQPDGWEISGFEKGDAGVYVNLKDTFTKPVFKIE 72
DB 72 LQDDVHSDEDI-----QLDSEDDSDTEAVQ-AQVVDKLAKDT--KSEQSLD 115

QY 73 EKKEENKFTFVSKKKONPVNHLQNES-----HRKEDLQR-----EEHS 114
DB 116 DELSEMDTKTVSLKLLKNEFVRSQVYSIIADTLLHRSNEVANANTKNSDDEHS 175

QY 115 QKSDTKDVTATVLDKNISSKSTT 139
DB 176 SKRKTKKKSITDFPKKQKNEDTT 200

RESULT 17
MDNL_HUMAN
ID MDNL_HUMAN STANDARD; PRT; 5596 AA.
AC Q9NU22; O15019;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Midasin (MDAS-containing protein).
GN MDNL OR KIAA0301.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX PubMed=12102729;
RA Garbarino J.E., Gibbons I.R.;
RT "Expression and genomic analysis of midasin, a novel and highly
RL conserved AAA protein distantly related to dynein.";
RN BMC Genomics 3:18-18(2002).
RP [2]
RN SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.
RC TISSUE=Testis;
RA Tracey A.;

```

---

```

RL Submitted (APR-2000) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE OF 3550-5596 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC -!- FUNCTION: May function as a nuclear chaperone and be involved in
CC the assembly/disassembly of macromolecular complexes in the
CC nucleus.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF503925; AAM77722.1; -.
CC EMBL; AL096678; CAB86660.1; -.
CC EMBL; AL096678; CAB86661.1; -.
CC EMBL; AB002299; BAA20761.1; -.
CC Genew; HGNC:18302; MDN1.
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0016887; F:ATPase activity; NAS.
CC GO; GO:0003754; F:chaperone activity; NAS.
CC GO; GO:0006461; P:protein complex assembly; NAS.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR002035; VWF_A.
CC Pfam; PF00004; AAA; 1.
CC SMART; SM00382; AAA; 7.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS50234; VWFA; 1.
KW Chaperone; ATP-binding; Repeat; Nuclear protein.
FT NP_BIND 329 336 ATP (POTENTIAL).
FT NP_BIND 677 684 ATP (POTENTIAL).
FT NP_BIND 1084 1091 ATP (POTENTIAL).
FT NP_BIND 1390 1397 ATP (POTENTIAL).
FT NP_BIND 1753 1760 ATP (POTENTIAL).
FT NP_BIND 2066 2073 ATP (POTENTIAL).
FT DOMAIN 3566 3573 POLY-GLU.
FT DOMAIN 4784 4791 POLY-GLU.
FT DOMAIN 5008 5013 POLY-GLU.
FT DOMAIN 5182 5187 VWFA.
FT DOMAIN 5384 5387 VWFA.
FT CONFLICT 2287 2312 RLFLSMDPVHGDISRAMNRLGIYI -> S (IN REF. 2).
SQ SEQUENCE 5596 AA; 632802 MW; 586C62616A1F96D4 CRC64;

Query Match 11.2%; Score 84; DB 1; Length 5596;
Best Local Similarity 23.2%; Pred. No. 2.7e+02;
Matches 29; Conservative 27; Mismatches 61; Indels 8; Gaps 2;

QY 1 HRVTVTIQKEMSSSTIVSEDFILPVYKGELEK-----YQFDGWEISGFEKGDAGY 54
DB 4632 HRSTAKLSVLAAQVFTLAQKGFCLPKFEMDSAGEGATEFHVDYGGGIGEGEGMKDVS- 4690
QY 55 VINLSKDTFIRPVFKKIEEKEENKPTFDVSKKDNPNQVNHSQLNSHRKEDLQREHS 114
DB 4691 -DQIGNERQVEDTFQKGQEKDQKDPDSKSDIKGEDNAIEMSEBDFDKVHGDGELEQEDD 4749
QY 115 QKSDS 119
DB 4750 EKSDS 4754

```

## RESULT 18

```

U2R1_HUMAN
ID U2R1_HUMAN STANDARD; PRT; 479 AA.
AC Q15695; Q13570;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit
DE related protein 1.
GN U2AF1-RS1 OR U2AF1RS1 OR U2AFBPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96163878; PubMed=8586425;
RA Kitagawa K., Wang X., Hatada I., Yamaoka T., Nojima H.,
RA Inazawa J., Abe T., Mitsuura K., Oshimura M., Murata A., Monden M.,
RA Mukai T.;
RT "Isolation and mapping of human homologues of an imprinted mouse gene
RT U2af1-rs1.";
RL Genomics 30:257-263 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovarian carcinoma;
RX MEDLINE=96212931; PubMed=8630064;
RA Pearsall R.S., Shibata H., Brozowska A., Yoshino K., Okuda K.,
RA Dejong P.J., Plass C., Chapman V.M., Hayashizaki Y., Held W.A.;
RT "Absence of imprinting in U2AFBPL, a human homologue of the imprinted
RT mouse gene U2afp-rs1.";
RL Biochem. Biophys. Res. Commun. 222:171-177 (1996).
CC -!- SURCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -!- SIMILARITY: Contains 2 C3H1-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D49676; BAA08532.1; -
CC EMBL; U51224; AAA98669.1; -
CC MIM; 601079; -
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0003723; F:RNA binding activity; NAS.
CC InterPro; IPR000504; RNA rec mot.
CC InterPro; IPR000571; Znf_CCH.
CC Pfam; PF00076; rrm; 1.
CC Pfam; PF00642; zf-CCH; 2.
CC SMART; SM00360; RRM; 1.
CC SMART; SM00356; Znf_C3H1; 2.
CC PROSITE; PS0102; RRM; 1.
CC PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Zinc-finger; Repeat.
FT DOMAIN 203 309 RNA-BINDING (RRM).
SQ SEQUENCE 479 AA; 57643 MW; 96F326694BD4E7C0 CRC64;

```

```

Query Match 11.2%; Score 83.5; DB 1; Length 479;
Best Local Similarity 21.7%; Pred. No. 22;
Matches 26; Conservative 25; Mismatches 40; Indels 29; Gaps 4;
QY 50 KDAGYVNLNKDTFKIPKFKTEKK-----EENKPTFDVSKKDNQ 93
Db :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
44 RDSGLSQEEEDTFIEE--QLEEEKLLERERLHEEWLLREQKAQEEFRKKKEEA 101
QY 94 VN-----HSQNLNESHKEDQLREHSQKSDSTKDVATVLD--KNNISSKSTTNP 142
Db :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
102 KKWLEEQERKLKEQWKEQKEREEREQKQKEKEEAQKMLDQAEADLENSTTQNP 161

```

```

RESULT 19
IF31_HUMAN
ID IF31_HUMAN STANDARD; PRT; 258 AA.
AC O75822; Q98UD2; Q98Q2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Eukaryotic translation initiation factor 3 subunit 1 (eIF-3 alpha)
DE (eIF3 p35) (eIF3j).
GN EIF3S1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=99041954; PubMed=9822859;
RA Block K.L., Vornlocher H.-P., Hershey J.W.B.;
RT "Characterization of cDNAs encoding the p44 and p35 subunits of human
RT translation initiation factor eIF3.";
RL J. Biol. Chem. 273:31901-31908 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovarian carcinoma;
RX MEDLINE=2238257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buétow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Rouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA.
CC -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 12 DIFFERENT SUBUNITS (BY
CC similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U97670; AAC78729.1; -
CC EMBL; AK023388; BAB14555.1; -

```



RT "Molecular analysis and expression of the lipase of Staphylococcus  
 RL epidermidis.";  
 RN J. Gen. Microbiol. 139:267-277(1993).  
 RP [2]  
 RC SEQUENCE FROM N.A.  
 RA STRAIN-ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
 RA Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
 CC fatty acid anion.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS  
 CC NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE  
 CC CONVERSION).  
 CC -1- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M95577; AAA19729.1; -;  
 DR EMBL; AE016744; AAO03878.1; -;  
 DR PIR; A47705; A47705.  
 DR InterPro; IPR005877; Gpos\_Y5IRK.  
 DR InterPro; IPR000734; Lipase.  
 DR InterPro; IPR000379; Ser esters site.  
 DR Pfam; PF04650; Y5IRK signal; 1.  
 DR TIGRFAMs; TIGR01168; Y5IRK signal; 1.  
 DR PROSITE; PS00120; LIPASE\_SIG; 1.  
 KW Hydrolase; Lipid degradation; Zymogen; Signal; Complete proteome.  
 FT SIGNAL 1 35 POTENTIAL.  
 FT PROPEP 36 302 REMOVED IN THE MATURE FORM.  
 FT CHAIN 303 688 LIPASE.  
 FT ACT\_SITE 418 418 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 648 648 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CONFLICT 96 96 W -> L (IN REF. 1).  
 FT CONFLICT 120 120 E -> G (IN REF. 1).  
 SQ SEQUENCE 688 AA; 77343 MW; 6C95DB3A78AF86F6 CRC64;  
 Query Match 11.1%; Score 83; DB 1; Length 688;  
 Best Local Similarity 24.5%; Pred. No. 35;  
 Matches 35; Conservative 28; Mismatches 60; Indels 20; Gaps 6;  
 Qy 6 TIONGKMSSTIVSEEDFILPYKGELEKGYQFDGWEISGFE-GKKDAGYVINLSKDTFI 64  
 Db 79 SITENESLHNETPKNEDWI-----QQQKDSQNDNKSESVEQNKENAFQVNHSEE--- 129  
 Qy 65 KPVFKKIE-EKKKEENKPTFDVSKKDNQV--NHSQLNESHKEDLQREHSHQKSDSTK 121  
 Db 130 KPQQQVELEKHASENNOTLHSAQAQSNEDVKTPEQLDNTAAQKQEDSKENLSKQDTQS 189  
 Qy 122 DVTATVLDKNISSSTNNPNK 144  
 Db 190 SKTTDLL-----RATAQNSK 205  
 RESULT 22  
 ADSV\_HUMAN  
 ID ADSV\_HUMAN STANDARD; PRT; 715 AA.  
 AC Q9Y6U3; Q8MU97;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Adseverin (Scinderin).  
 GN SCIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Pahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [2]  
 RP SEQUENCE OF 1-527 FROM N.A.  
 RN Kalicki J., Smith-Craig R.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE OF 248-715 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Isoqai T., Oca T., Hayashi K., Sugiyama T., Oteuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Ca(2+)-dependent actin filament-severing protein that is  
 CC presumed to have a regulatory function in exocytosis by affecting  
 CC the organization of the microfilament network underneath the  
 CC plasma membrane. In vitro, also has barbed end capping and  
 CC nucleating activities in the presence of Ca(2+).  
 CC -1- SIMILARITY: BELONGS TO THE VILIN/GELSOLIN FAMILY.  
 CC -1- SIMILARITY: Contains 6 gelsolin-like repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; BC021090; AAH21090.1; -;  
 DR EMBL; AC005281; AAD15423.1; -;  
 DR EMBL; AK027778; BAB55361.1; -;  
 DR HSSP; P02640; 2VIL.  
 DR InterPro; IPR001974; Gelsolin.  
 DR Pfam; PF00626; Gelsolin; 6.  
 DR PRINTS; PR00597; GELSOLIN.  
 DR SMART; SM00262; GEL; 6.  
 KW Cytoskeleton; Actin-binding; Repeat; Calcium; Capping protein.  
 FT DOMAIN 1 363 ACTIN-SEVERING (POTENTIAL).  
 FT DOMAIN 364 715 CA(2+)-DEPENDENT ACTIN BINDING.  
 FT REPEAT 27 76 GELSOLIN-LIKE 1.  
 FT REPEAT 148 188 GELSOLIN-LIKE 2.  
 FT REPEAT 265 307 GELSOLIN-LIKE 3.  
 FT REPEAT 398 451 GELSOLIN-LIKE 4.  
 FT REPEAT 523 564 GELSOLIN-LIKE 5.  
 FT REPEAT 626 668 GELSOLIN-LIKE 6.  
 FT SITE 112 119 POLYPHOSPHOINOSITIDE BINDING (BY

FT SITE 138 146 SIMILARITY).  
 FT POLYPHOSPHONOSITIDE BINDING (BY  
 FT SIMILARITY).  
 FT CONFLICT 61 61 R -> H (IN REF. 2).  
 SQ SEQUENCE 715 AA; 80508 MW; 45PBE42CBCEFDDB0 CRC64;  
 Query Match 11.1%; Score 83; DB 1; Length 715;  
 Best Local Similarity 23.1%; Pred. No. 36;  
 Matches 34; Conservative 30; Mismatches 47; Indels 36; Gaps 5;  
 QY 2 RVTVTIQKEMSSITVSEEDFTLPVYKGELEKGYQFDCWEISGEGK-----KDAGYV 55  
 Db 262 RVTVAENPFMSMILLSECFILD--HGAAKQIFVWKGKDANPQRKAMKTAEFLOQ 319  
 QY 56 INLSKDTFIK-----VFFKK-----IEEKEEENKPTFDVSKKK 89  
 Db 320 MYVSKNTQIQLVPEGGETIFKOFFKDWDRKQSDGFGVKYVTEKVAIKQIPFDASKLH 379  
 QY 90 DNPQV--NHSQLNESHKEDLOREHS 114  
 Db 380 SSPQMAAQNMDVDDSGKVEIWRVENN 406

RESULT 23  
 GLNA\_PIRAB STANDARD; PRT; 439 AA.  
 AC Q9UY99;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
 GLNA OR PYRAB16090 OR PAB1292.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GES / Orsay;  
 RX PubMed=12622808;  
 RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,  
 RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,  
 RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;  
 RT "An integrated analysis of the genome of the hyperthermophilic  
 archaeon Pyrococcus abyssi.".  
 RL Mol. Microbiol. 47:1495-1512(2003).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
 L-glutamine. LOCATION: Cytoplasmic (Potential).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ248288; CAB50513.1; -.  
 DR FIR; C75009; C75009.  
 DR HSSP; P06201; 1LGR.  
 DR InterPro; IPR001691; Gln synth.  
 DR InterPro; IPR004809; GlnA.  
 DR InterPro; IPR001637; GlnA adenyltn.  
 DR Pfam; PF00120; gln-synt; 1.  
 DR Pfam; PF03951; gln-synt N; 1.  
 DR ProDom; PD001057; Gln synt C; 1.  
 DR TIGRFAMs; TIGR00653; GlnA; 1.  
 DR PROSITE; PS00180; GLNA.1; 1.  
 DR PROSITE; PS00181; GLNA-ATP; 1.  
 KW Ligase; Complete proteome.  
 FT BINDING 358 358 AMP (UNDER CONDITIONS OF ABUNDANT

FT SEQUENCE 439 AA; 49742 MW; 62CCFD3970A98AF0 CRC64;  
 Query Match 11.0%; Score 82.5; DB 1; Length 439;  
 Best Local Similarity 17.8%; Pred. No. 24;  
 Matches 28; Conservative 24; Mismatches 36; Indels 69; Gaps 4;  
 QY 32 LEKGYQFDCWEISGEGKDKAGYVINLSKDTFI-----KP----- 66  
 Db 42 IEDGISFDGSSVPGQIGIEDSLIFKADPTTYVEPNDVNARVYGYIYKDGKPYGADPRG 101  
 QY 67 VPKTEEKKEEN-----KPTDVSKKKN----- 91  
 Db 102 VLKRVIEKLAEMGKIKAYIGPEFYLFKKGSGWELEIPDVGYFDILTLDKAKOIKREIA 161  
 QY 92 -----PQVNHSQLNESHKEDLOREHSQKSDS 119  
 Db 162 EYMPFGLVPEVLHVEVGKQAHEIDFRYDEALKTADN 198

RESULT 24  
 DNLI\_CABEL STANDARD; PRT; 847 AA.  
 AC Q27474;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA ligase (EC 6.5.1.1) (Polydeoxyribonucleotide synthase [ATP]).  
 GN LIG-1 OR C29A12.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Wilkinson J.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA  
 CC RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA.  
 CC -!- CATALYTIC ACTIVITY: ATP + {deoxyribonucleotide}(N) +  
 {deoxyribonucleotide}(M) = AMP + diphosphate +  
 {deoxyribonucleotide}(N+M).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z73970; CAA98242.1; -.  
 DR FIR; T19544; T19544  
 DR WormPep; C29A12.3; CE05328.  
 DR InterPro; IPR000977; DNA ligase.  
 DR Pfam; PF01068; DNA ligase; 1.  
 DR Pfam; PF04679; DNA ligase A C; 1.  
 DR Pfam; PF04675; DNA ligase A N; 1.  
 DR TIGRFAMs; TIGR00574; dnli; 1.  
 DR PROSITE; PS00697; DNA\_LIGASE\_A1; 1.  
 DR PROSITE; PS00333; DNA\_LIGASE\_A2; 1.  
 DR PROSITE; PS0160; DNA\_LIGASE\_A3; 1.  
 KW DNA repair; DNA replication; DNA recombination; Cell division; Ligase;  
 KW ATP-binding.  
 FT BINDING 363 363 AMP (BY SIMILARITY).  
 SQ SEQUENCE 847 AA; 94501 MW; A32B7D279F91002F CRC64;

Query Match 11.0%; Score 82.5; DB 1; Length 847;  
 Best Local Similarity 26.3%; Pred. No. 48;  
 Matches 40; Conservative 23; Mismatches 60; Indels 29; Gaps 7;  
 QY 7 IQNGKEMSTIVSEEDFTLPVYKGE-----LEKGYQFDCWEISGEGKDKAGYVINL- 58



```

DB 671 IRDDKNSDDATSSQ--VLEMYKNOEAFANQIKEXADAVD--EDDEFKEDEEBELNMT 726
QY 59 -----SKDTFKPKVFKKIEEKEENKPTFDVSKK--DNPQVNHSLQSHRKHEDLQR 110
DB 727 NVSESSKENPVK-----EIKKSTPK--SVSPKFKFKPPVXSSPVNKSPPVKSSPIK 777
QY 111 EEHSQKSDSTKDVATVLDKNNISSKSTNNP 142
DB 778 KEAEKKGPVASIFSSSTKQNEKDVKVESPP 809

RESULT 25
ARS2 DROME STANDARD; PRT; 943 AA.
AC Q9V9K7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arsenite-resistance protein 2 homolog.
GN CG7843.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Aehburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe W., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL SEQUENCE 287:2185-2195(2000).
RN [2]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RC STRAIN=Berkeley;
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

```

```

RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q9V9K7-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q9V9K7-2; Sequence=VSP 000327;
CC -!- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003784; AAM68343.1; -
DR EMBL; AE003784; AAM68345.1; -
DR FlyBase; FBGN0033062; CG7843.
DR Pfam; PF04959; ARS2; 1.
KW Hypothetical protein; Alternative splicing.
FT VARSPLIC 47 50 Missing (in isoform Short).
FT SEQUENCE 943 AA; 107221 MW; 0C1AF05E02E8AB0B CRC64;
SQ
Query Match 11.0%; Score 82.5; DB 1; Length 943;
Best Local Similarity 25.6%; Pred.No.53;
Matches 32; Conservative 21; Mismatches 61; Indels 11; Gaps 3;
QY 7 IQNGKMSSTI-----VSEDFILPVVKGLEKGYQFDGWISGFEKGKDGAVVNLK 60
DB 291 MQSVKEVKTINSKPEEMSEADPVSTQKPKVPVNSDGENWDDDAENSAPKELAESK 350
QY 61 DTFIKPVKITEE---KKEENKPTFDVSKKKNPQVNHSLQSHRKHEDLQREHSQKS 117
DB 351 DSDSKPEDKQLNKKTKRKKNSSDDSSSESSSSDEEKLKEDYDVEDGLRAE--QKT 408
QY 118 DSTKD 122
DB 409 EAERD 413
RESULT 26
MAPB RAT STANDARD; PRT; 2459 AA.
AC P15205; Q62958; Q9ER21; Q9QW92;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
DE light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
RT associated protein (MAP1B)-encoding cDNA."
RL Gene 172:307-308(1996).

```



RESULT 30

```

GLNA_PYWRO STANDARD; PRT; 439 AA.
ID ID GLNA_PYWRO PRT; 439 AA.
P36687;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DT Glutamine synthetase I (EC 6.3.1.2) (Glutamate--ammonia ligase) (GSI).
DE GLNA.
DN
OS Pyrococcus woesei.
OS Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OC NCBI_TaxID=2262;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=93259940; PubMed=8098326;
RA Tiboni O., Cammarano P., Sanangelantoni A.M.;
RT "Cloning and sequencing of the gene encoding glutamine synthetase I
RT from the archaeum Pyrococcus woesei: anomalous phylogenies inferred
RT from analysis of archaeal and bacterial glutamine synthetase I
RT sequences.";
RL J. Bacteriol. 175:2961-2969 (1993).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X60161; CAA42730.1; -.
DR PIR; A36911; A36911.
DR HSSP; P06201; ILGR.
DR InterPro; IPR001691; GLN synth.
DR InterPro; IPR004809; GlnA.
DR InterPro; IPR001637; GlnA_adenyln.
DR Pfam; PF00120; gln-synt; 1.
DR Pfam; PF03951; gln-synt N; 1.
DR ProDom; PD001057; Gln_synt C; 1.
DR TIGRFAMs; TIGR00653; GlnA; 1.
DR PROSITE; PS00180; GlnA_1; 1.
DR PROSITE; PS00181; GlnA_ATP; 1.
KW L-gase.
FT FT BINDING
FT FT BINDING
SQ SEQUENCE 439 AA; 50066 MW; C503B43B790BD25C CRC64;
Query Match 11.0%; Score 82; DB 1; Length 439;
Best Local Similarity 18.1%; Pred.No. 26;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;
QY 35 GYDFDGEWISGFEGKDGAGVYINLSKDTFI-----KP-----VFK 69
DB 45 GISEFGSSVFGQIEDSLVFRADPTIYVEVPDWNVARYVGYIKNKYGDPRGILK 104
QY 70 KIEKKEEN-----KPTFDVSKKDN----- 91
DB 105 RALEELEKEGYKAVIGPEPEFYLFKNGTWELEIPDVGYYFDILTDKARDIRIEAYM 164
QY 92 -----PQVNHSQLNESHKEDLOREHSQSD---STKDVATVLDKNNI 133
DB 165 PSFGLIEVLHHEVGKAQHEIDFRYDEALKTADNIVSFKYITTKAVAEHGL 215
RESULT 31
BPBP_BACSU STANDARD; PRT; 914 AA.
ID ID BPBP_BACSU
AC P39793;
DT 01-FEB-1995 (Rel. 31, Created)

```



```
SQ SEQUENCE 949 AA; 105961 MW; FA8969B0C64B3278 CRC64;
Query Match 11.0%; Score 82; DB 1; Length 949;
Best Local Similarity 28.7%; Pred. No. 58;
Matches 29; Conservative 20; Mismatches 20; Indels 32; Gaps 5;
QY 60 KDTFKVPKKEEKEENKTFDVSKKD-----NPQ-VNHSLNES 102
DB 98 EETKAPKPKTKKEKKEAP---IHKKEIEIVNTFENOTPLVENTPKAVSHSQIEKA 154
QY 103 HRKEDLOREHSOKSDSTKDVATVLDKNNISKSTNNPN 143
DB 155 --KQKLEIQKREA-----LNKLTQNTNTNNAN 183
RESULT 33
LMD1_HUMAN
ID LMD1_HUMAN STANDARD; PRT; 572 AA.
AC P29536;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Leiomodlin 1 (Leiomodlin, muscle form) (64 kDa autoantigen D1) (64 kDa
DE autoantigen ID) (64 kDa autoantigen ID3) (Thyroid-associated
DE ophthalmopathy autoantigen) (Smooth muscle leiomodlin) (SM-Lmod).
GN LMOD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=91225220; PubMed=2026759;
RA Dong Q., Ludgate M., Vaseart G.;
RT "Cloning and sequencing of a novel 64-kDa autoantigen recognized by
RT patients with autoimmune thyroid disease.";
RL J. Clin. Endocrinol. Metab. 72:1375-1381(1991).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99451105; PubMed=10520227;
RA Conley C.A., Fowler V.M.;
RT "Localization of the human 64kD autoantigen D1 to myofibrils in a
RT subset of extraocular muscle fibers.";
RL Curr. Eye Res. 19:313-322(1999).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=21248187; PubMed=11350761;
RA Conley C.A.;
RT "Leiomodlin and tropomodulin in smooth muscle.";
RL Am. J. Physiol. 280:C1645-C1656(2001).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=21218919; PubMed=11318603;
RA Conley C.A., Fritz-Six K.L., Almenar-Queralt A., Fowler V.M.;
RT "Leiomodlins: larger members of the tropomodulin (Tmod) gene family.";
RL Genomics 73:127-139(2001).
RN [5]
RP SUBCELLULAR LOCATION: CYTOSKELETON.
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE (HEART, SKELETAL MUSCLE, COLON
CC AND SMALL INTESTINE), A SUBSET OF STRIATED MUSCLE FIBERS, AND AT
CC LOW LEVEL IN THYROID.
CC -1- DISEASE: RECOGNIZED BY PATIENTS WITH AUTOIMMUNE THYROID DISEASE.
CC -1- SIMILARITY: BELONGS TO THE TROPOMODULIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; X54162; CAA38101.1; -.
```

```
DR PIR; S18732; S18732.
DR Genew; HGNC:6647; LMOD1.
DR MIM; 602715; -.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0009405; P:pathogenesis; TAS.
DR InterPro; IPR004934; Tropomodulin.
DR InterPro; IPR003124; WH2
DR Pfam; PF03250; Tropomodulin; 1.
DR Pfam; PF02205; WH2; 1.
DR SMART; SM00246; WH2; 1.
DR Antigen; Repeat; Cytoskeleton.
KW DOMAIN 137 265 8 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 137 152 1.
FT REPEAT 153 168 2.
FT REPEAT 169 184 3.
FT REPEAT 185 200 4.
FT REPEAT 201 216 5.
FT REPEAT 217 232 6.
FT REPEAT 233 248 7.
FT REPEAT 249 265 8.
FT DOMAIN 480 499 5 X 4 AA APPROXIMATE TANDEM REPEATS.
SQ SEQUENCE 572 AA; 63737 MW; D4B42F8E0523DE94 CRC64;
Query Match 10.9%; Score 81.5; DB 1; Length 572;
Best Local Similarity 22.8%; Pred. No. 37;
Matches 38; Conservative 21; Mismatches 57; Indels 51; Gaps 6;
QY 29 KGELEKGYQFDGWEISGFEGK--KDAGYVINLSKDTFIKVPFK-----IEEK 74
DB 99 RGLKKSFSRDRDEAGGKSGEKPKEKIIRGIDKGRVRAAVDKKEAGKDGREERAVATK 158
QY 75 KEENK-----PTFDVSKKDKNPQVNHSLNESHRKE----- 106
DB 159 KEEKKGGDRNTGLSRDKDKKEEMKEVAKEDDEKVGGERNTDTRKEGEMKRAAGNT 218
QY 107 DLQREHSQK-----SDSTKDVAT-----VLDKNNISSKSTNNPNK 144
DB 219 DMKEDKVKRGNTGNTDKDDEKVKKNPLHEKEAKDDSKTKTPEK 265
RESULT 34
DRSL_YEAST
ID DRSL_YEAST STANDARD; PRT; 752 AA.
AC P32892;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable ATP-dependent RNA helicase DRS1.
GN DRS1 OR YLL008W OR L1345.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93087480; PubMed=1454790;
RA Ripmaster T.L., Vaughn G.P., Woolford J.L. Jr.;
RT "A putative ATP-dependent RNA helicase involved in Saccharomyces
RT cerevisiae ribosome assembly.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / FY23;
RX MEDLINE=96405918; PubMed=8810043;
RA Miosga T., Zimmermann F.K.;
RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT a 4.7 kb fragment of chromosome XII including an open reading frame
RT homologous to the human cystic fibrosis transmembrane conductance
RT regulator protein CFTR.";
RL Yeast 12:693-708(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
```

```

RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansoorge W.,
RA Benes V., Brueckner M., Dilius H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Mioega T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Honeisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:87-90(1997).
CC -!- FUNCTION: PROBABLE HELICASE INVOLVED IN RIBOSOME ASSEMBLY.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L00683; AAA34666.1; -
DR EMBL; X91488; CA62783.1; -
DR EMBL; Z73113; CA97452.1; -
DR PIR; S64750; S64750.
DR HSSP; Q58083; 1HV8.
DR SGD; S0003931; DRS1.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000629; DEAD_box.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
KW ATP-binding; RNA-binding; Helicase; Nuclear protein.
FT DOMAIN 170 190 POLY-GLU.
FT NP BIND 275 282 ATP (POTENTIAL).
FT SITE 385 388 DEAD BOX.
FT CONFLICT 1 68 MVGTGKYNLDVFPTISDSDDVILSDSDDEKVEAKKT
FT KRSRGKRNKKYVSGDNLDEVDHEDLD -> MTKSLRLRL
FT RSRGRVLRTRKRLVKGITSMFMPKWT (IN REF. 1).
SQ SEQUENCE 752 AA; 84843 MW; 60747607A6E5E4A8 CRC64;
Query Match 10.9%; Score 81.5; DB 1; Length 752;
Best Local Similarity 21.0%; Pred. No. 50;
Matches 30; Conservative 26; Mismatches 48; Indels 39; Gaps 6;
Qy 18 VSEDFILPVYKLEKGYQFD-----QWEI--SGFEKGDAGYVINLSKDTFI 64
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 VSEGNLDEVDHEDLDAGFKFDLADDTTSNFQGNFLAEGSNKDDAEAFV--KKDVL 110
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 65 KPVFKK-----IEKKEENKPTFDVSKKDN-----PQVHNSQLNE 101
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 DKIIIRKGLVKMAHIDSKQEETEKE-KVEKENDSDDELDAMDFGGMNNGENQS 169
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 102 SHRKEDLQREHSQKSDSTKDV 124
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 EEESEEEKEEEEEEEQEEEMT 192
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 35
DACA BACSU
ID DACA BACSU STANDARD; PRT; 443 AA.
AC P08750;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE D-alanyl-D-alanine carboxypeptidase precursor (EC 3.4.16.4) (DP-
DE peptidase) (DD-carboxypeptidase) (CPase) (PBPs).
GN DACA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RL DNA Res. 1:1-14(1994).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takanashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RL subtilis."
RL Nature 390:249-256(1997).
RN [3]
RN SEQUENCE OF 32-102.
RX MEDLINE=90182289; PubMed=6768745;
RA Waxman D.J., Strominger J.L.;
RT "Sequence of active site peptides from the penicillin-sensitive D-
RT alanine carboxypeptidase of Bacillus subtilis. Mechanism of
RT penicillin action and sequence homology to beta-lactamases."
RL J. Biol. Chem. 255:3964-3976(1980).
RN [4]
RN SEQUENCE OF 103-443 FROM N.A.
RX MEDLINE=86250602; PubMed=3087956;
RA Todd J.A., Roberts A.N., Johnstone K., Piggot P.J., Winter G.,
RA Ellar D.J.;
RT "Reduced heat resistance of mutant spores after cloning and
RT mutagenesis of the Bacillus subtilis gene encoding penicillin-binding
RT protein 5."
RL J. Bacteriol. 167:257-264(1986).
RN [5]
RN SEQUENCE OF 414-443.
RX MEDLINE=81117303; PubMed=6780559;
RA Waxman D.J., Strominger J.L.;
RT "Primary structure of the COOH-terminal membranous segment of a
RT penicillin-sensitive enzyme purified from two Bacilli."
RL J. Biol. Chem. 256:2067-2077(1981).

```





CC ----- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59720; CAA42405.1; -  
 DR EMBL; X53998; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; S74279; S74279.  
 DR SGD; S0000566; YCL061C.  
 DR GO; GO:0006347; P:chromatin silencing at HML and HMR (sensu S. . .); IGI.  
 DR GO; GO:0006348; P:chromatin silencing at telomere; IGI.  
 DR GO; GO:0000076; P:DNA replication checkpoint; IGI.  
 KW Hypothetical protein.  
 FT CONFLICT 505 505 L -> V (IN REF. 3).  
 FT CONFLICT 567 567 MISSING (IN REF. 3).  
 SQ SEQUENCE 853 AA; 97946 MW; 16E09FCC0BF248D1 CRC64;  
 Query Match 10.8%; Score 81; DB 1; Length 853;  
 Best Local Similarity 21.7%; Pred. No. 62;  
 Matches 34; Conservative 21; Mismatches 68; Indels 34; Gaps 5;  
 QY 11 KEMSTIVSEEDFILPVY-----KGELEKGYQFDGWEISGFEG-----KKD 51  
 DB 293 KEKREKLEENDFQNAHDSGSDSGSESGFALSGNEIADYESSGSENDNRSESEKED 342  
 QY 52 AGYVNLKDTPIKPFVKIKIEKKEENKPTFDVSKKXNDPOVNHQSOLNESHKEDLORE 111  
 DB 343 DEILKQKKSHVKKHIIINESDSTEVEAKP-----KEKADESIPKRIAINLGHYGDNI--- 395  
 QY 112 EHSQKSDTKDVTATVLDKNNI-----SSKSTNNPNK 144  
 DB 396 ----GEDTDKFOETNVLPTQNIIEVMAERTNIENEVK 428  
 RESULT 38  
 YAE6 SCHPO  
 ID YAE6 SCHPO STANDARD; PRT; 1325 AA.  
 AC Q09847;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C23D3.06C in chromosome I.  
 GN SPAC23D3.06C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne M., Stewart A.,  
 RA Brooks J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Collins M., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Muehlbauer H.,  
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.K., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe."  
 RL Nature 415:871-880(2002).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z64354; CAA91241.1; -  
 DR PIR; S62497; S62497.  
 DR GenedB SPombe; SPAC23D3.06C; -  
 DR Pfam; PF03093; Nucleoporin\_FG; 10.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1325 AA; 145776 MW; EEPD952FB7F0E6BD CRC64;  
 Query Match 10.8%; Score 81; DB 1; Length 1325;  
 Best Local Similarity 23.3%; Pred. No. 99;  
 Matches 35; Conservative 19; Mismatches 40; Indels 56; Gaps 6;  
 QY 41 WEISGFEGKKDAGYVNLKSD-----TFIKPVFKIEKKEEN---KP-TFDV 85  
 DB 685 WGASTFOSKQPSFSGFLTLDDKSNTPGKNFSIFGKTAETQVEQKPPNNVLTKPFSFAP 744  
 QY 86 SKK-----KDNQVNHQSOLNESHKEDLQREHSOKSDSKDVTVA 125  
 DB 745 SDKSMFAANIPISAGEGLDQKTSKALPSTGTLKJSEND-----NKAESNETKGFNT 797  
 QY 126 TVLDKNNISSK-----TTNN 141  
 DB 798 TIAKQNDKSKSEKSGKASVANNMALKNSINN 827  
 RESULT 39  
 UN89 CAEEL  
 ID UN89 CAEEL STANDARD; PRT; 6632 AA.  
 AC Q01761; Q17362;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).  
 GN UNC-89 OR C09D1.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=96180278; PubMed=8603916;  
 RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;  
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line  
 RT assembly, encodes a giant modular protein composed of Ig and signal  
 RT transduction domains."  
 RL J. Cell Biol. 132:835-848(1996).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RC STRAIN=Bristol N2;  
 RA Du Z., Le T.T., Wilson R.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN REVISIONS.  
 RP Waterston R.;  
 RA Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 25.4955 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTVTIQKMSSTIVSE.....ATVLKKNISSKSTNNPNK 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*\*

- 1: sp\_archaea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mhc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phase:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_rodent:\*\*
- 12: sp\_virus:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_unclassified:\*\*
- 15: sp\_virus:\*\*
- 16: sp\_bacteriap:\*\*
- 17: sp\_archheap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	748	100.0	2119	2 Q9AHT5	Q9Aht5 streptococc
2	748	100.0	2140	16 Q97RY6	Q97ry6 streptococc
3	745	99.6	2144	16 Q8DQF7	Q8dqp7 streptococc
4	744	99.5	2144	2 Q9S4M8	Q9s4m8 streptococc
5	112	15.0	361	5 Q95P15	Q95p15 plasmodium
6	110	14.7	346	5 Q9U0G0	Q9u0g0 plasmodium
7	110	14.7	379	5 Q25705	Q25705 plasmodium
8	110	14.7	775	16 Q8CPK8	Q8cpk8 staphylococ
9	109	14.6	3008	5 Q81436	Q81436 plasmodium
10	108.5	14.5	600	5 Q77355	Q77355 plasmodium
11	108.5	14.5	1038	13 Q90784	Q90784 gallus gall
12	108	14.4	354	5 Q25995	Q25995 plasmodium
13	108	14.4	354	5 Q81J55	Q81j55 plasmodium
14	107	14.3	829	5 Q815F3	Q815f3 plasmodium
15	106.5	14.2	470	10 Q9FJK9	Q9fjk9 arabidopsis
16	104.5	14.0	379	5 Q9U6C4	Q9u6c4 plasmodium

17	104.5	14.0	380	5	Q26019	Q26019 plasmodium
18	104	13.9	951	5	Q96229	Q96229 plasmodium
19	103.5	13.8	325	5	O44016	O44016 dictyosteli
20	103.5	13.8	379	5	Q25706	Q25706 plasmodium
21	102.5	13.7	2081	10	Q9LH98	Q9lh98 arabidopsis
22	102.5	13.6	382	5	Q9V7J0	Q9v7j0 drosophila
23	101.5	13.6	556	5	Q9V7I9	Q9v7i9 drosophila
24	101.5	13.6	785	5	Q9GQ82	Q9gq82 drosophila
25	101	13.5	329	5	Q9NFV9	Q9nf9v plasmodium
26	100	13.4	312	16	Q9PPL5	Q9ppl5 campylobact
27	100	13.4	1130	5	Q8IJZ4	Q8ijz4 plasmodium
28	99	13.2	211	5	Q9I488	Q9i488 caenorhabdi
29	99	13.2	1859	5	Q8IC37	Q8ic27 plasmodium
30	99	13.2	2563	5	Q8I3A0	Q8i3a0 plasmodium
31	98	13.1	157	5	Q9VQV0	Q9vgv0 drosophila
32	97.5	13.0	449	5	Q8IHW3	Q8ihw3 plasmodium
33	97.5	13.0	556	5	Q9S593	Q9s593 drosophila
34	97	13.0	558	5	Q8I5T1	Q8i5t1 plasmodium
35	97	13.0	988	10	Q9STN4	Q9stn4 arabidopsis
36	96	12.8	238	5	Q8I2Z6	Q8i2z6 plasmodium
37	96	12.8	531	5	Q8I426	Q8i426 plasmodium
38	96	12.8	1795	5	Q8IE35	Q8ie35 plasmodium
39	95.5	12.8	375	4	O14712	O14712 homo sapien
40	95.5	12.8	635	16	Q8ESX9	Q8esx9 streptococc
41	95.5	12.8	641	10	Q93ZQ2	Q93zq2 arabidopsis
42	95.5	12.8	644	10	Q9M1D2	Q9m1d2 arabidopsis
43	95	12.7	535	5	Q17595	Q17595 caenorhabdi
44	95	12.7	622	16	Q98QAL	Q98qal mycoplasma
45	95	12.7	3026	5	Q8ILS9	Q8ils9 plasmodium

#### ALIGNMENTS

#### RESULT 1

Q9AHT5 PRELIMINARY; PRT; 2119 AA.

AC Q9AHT5; 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Serine protease (Fragment).

GN PRTA.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_taxid=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=N4;

RX MEDLINE=21116976; Pubmed=11179332;

RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,

RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,

RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,

RA Langermann S., Johnson S., Koenig S.;

RT "Use of a Whole Genome Approach To Identify Vaccine Molecules

RT Affording Protection against Streptococcus pneumoniae Infection.";

RL Infect. Immun. 69:1593-1598(2001).

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY

CC AN AMIDE BOND (BY SIMILARITY).

CC EMBL; AF291699; AAK19159.1; -.

DR HSSP; P00782; 2SPT.

DR MEROPS; S08.064; -.

DR InterPro; IPR001899; Gram\_pos\_anchor.

DR InterPro; IPR006192; LPXTG.

DR InterPro; IPR003137; PA.

DR InterPro; IPR000209; Peptidase\_S8.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00746; Gram\_pos\_anchor; 1.

DR Pfam; PF02225; PA; 1.

DR Pfam; PF00082; Peptidase\_S8; 2.

DR PRINTS; PR00723; SUBTILISIN.

DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.

```
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Cell wall; Peptidoglycan-anchor; Protease.
FT NON_TER 1
SQ SEQUENCE 2119 AA; 238226 MW; 517F9B7F6B960A6A CRC64;

Query Match      100.0%; Score 748; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 9.4e-49;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKKDAGYVNLK 60
Db 1942 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKKDAGYVNLK 2001

QY 61 DTFIKPVFKIEEKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHHSOKSDST 120
Db 2002 DTFIKPVFKIEEKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHHSOKSDST 2061

QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
Db 2062 KDVTATVLDKNNISSKSTTNNPNK 2085

RESULT 2
Q97RY6 PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Serine protease, subtilase family.
GN SP0641.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=1146316;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umavam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506 (2001).
DR EMBL; AE007373; AAK74791.1; -.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Peptidase_S8.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRAME; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Protease; Complete proteome.
```

```
SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;

Query Match      100.0%; Score 748; DB 16; Length 2140;
Best Local Similarity 100.0%; Pred. No. 9.5e-49;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKKDAGYVNLK 60
Db 1963 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKKDAGYVNLK 2022

QY 61 DTFIKPVFKIEEKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHHSOKSDST 120
Db 2023 DTFIKPVFKIEEKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHHSOKSDST 2082

QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
Db 2083 KDVTATVLDKNNISSKSTTNNPNK 2106

RESULT 3
Q8DQP7 PRELIMINARY; PRT; 2144 AA.
AC Q8DQP7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN PrtA OR SPR0561.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., BURGESS S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McMahon S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717 (2001).
DR EMBL; AE008434; AAK9365.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match      99.6%; Score 745; DB 16; Length 2144;
Best Local Similarity 99.3%; Pred. No. 1.6e-48;
Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKKDAGYVNLK 60
Db 1967 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKKDAGYVNLK 2026

QY 61 DTFIKPVFKIEEKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHHSOKSDST 120
Db 2027 DTFIKPVFKIEEKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHHSOKSDST 2086

QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
Db 2087 KDVTATVLDKNNISSKSTTNNPNK 2110

RESULT 4
Q9S4M8 PRELIMINARY; PRT; 2144 AA.
AC Q9S4M8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
```

01-MAR-2003 (Tremblrel. 23, Last annotation update)  
Cell wall-associated serine proteinase PrtA precursor.  
PRTA.  
Streptococcus pneumoniae.  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
NCBI\_TaxID=1313;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=3 B;  
Bethe G., ten Thoren E., Bongarts R.J.M., Heinz H.-P., Zysk G.;  
"Cloning and sequencing of a novel surface protease of Streptococcus  
pneumoniae";  
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RL SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
CC -1-  
AN AMIDE BOND (BY SIMILARITY).  
DR EMBL; AF127143; AAD48399.1; -.  
DR HSP; P00782; 2SBI.  
DR MEROPS; S08.064; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR PRINTS; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PF00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Cell wall; Peptidoglycan-anchor; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 2144 CELL WALL-ASSOCIATED SERINE PROTEINASE  
PRTA.  
SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
  
Query Match 99.5%; Score 744; DB 2; Length 2144;  
Best Local Similarity 98.6%; Pred. No. 1.9e-48;  
Matches 142; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 HRVTVIQNGEMSTIVSEEDFILPVYKLEKGYQFDGWISGFEKGDAGVYINLSK 60  
Db 1967 HRVTVIQNGEMSTIVSEEDFILPVYKLEKGYQFDGWISGFEKGDAGVYINLSK 2026  
  
Qy 61 DTFIKPVFKIEEKEENKPTFDVSKKKNPQVNHSQLNESHKREDLQREHSQKSDST 120  
Db 2027 DTFIKPVFKIEEKEENKPTFDVSKKKNPQVNHSQLNESHKREDLQREHSQKSDST 2086  
  
Qy 121 KDVTATVLDKNNISSKSTNNPNK 144  
Db 2087 KDVTATVLDKNNISSKSTNNPNK 2110  
  
RESULT 5  
Q95PI5 ID Q95PI5 PRELIMINARY; PRT; 361 AA.  
AC Q95PI5  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE Merozoite surface protein 3 (fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVO;  
RA Hiseida H., Saul A., Long C.A., Miller L.H., Stowers A.W.;  
RT "Merozoite Surface Protein 3 and Protection Against Malaria in Aotus

Monkeys";  
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY041180; AAK94780.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 361 361  
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;  
  
Query Match 15.0%; Score 112; DB 5; Length 361;  
Best Local Similarity 23.9%; Pred. No. 0.57;  
Matches 37; Conservative 28; Mismatches 48; Indels 42; Gaps 6;  
  
Qy 28 YKGELEKGYQ-----FD---GWEISGF--EGKKGAG-----YVI 56  
Db 165 YAGKVKDYERAKNAYQKANQAVLKAKEASSYDILGWFEFGGVPEHKKEENMLSHLYVS 224  
Qy 57 NLSKDTFIKPVFKIEEKEENKPTFDVSKKKNPQVNHSQLNESHKREDLQREHSQK 116  
Db 225 SKKENISKENDVDLDE-KEEAEETEELKEEKEETEISEDEDEEEEEEEEEEE 283  
  
Qy 117 SDSTKQVATVLDKN-----NISKSTNN 141  
Db 284 NDKKKEQESQNNENNNDKQKDMQAQNLISKNNNN 318  
  
RESULT 6  
Q9U0G0 ID Q9U0G0 PRELIMINARY; PRT; 346 AA.  
AC Q9U0G0  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE Merozoite surface protein 3 (fragment).  
GN MSP3.  
OS Plasmodium reichenowi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5854;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20416497; PubMed=10960178;  
RA Okenu D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
Plasmodium reichenowi and Plasmodium falciparum";  
RL Mol. Biochem. Parasitol. 109:185-188 (2000).  
DR EMBL; AJ252286; CAB65754.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 346 346  
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAPA010 CRC64;  
  
Query Match 14.7%; Score 110; DB 5; Length 346;  
Best Local Similarity 26.2%; Pred. No. 0.78;  
Matches 34; Conservative 26; Mismatches 54; Indels 16; Gaps 5;  
  
Qy 27 VYKGELEKGYQFD-GWEISGF--EGKKGAG-----YVINLSKDTFIKPVFKIEEKEE 78  
Db 184 VLKAKEASSYNYILGWFEFGGVPEHKKEENMLSHLYSSKDKKENISKENDVDLDE-KEEE 242  
Qy 79 NKPTFDVSKKKNPQVNHSQLNESHKREDLQREHSQKSDSTKQVATV-----LDKN 131  
Db 243 AETGEQELKEEKEETEISEINEDEQEKEEKEEENNDKKEQAKESQNDQKEDMEAQ 302  
  
Qy 132 NISKSTNN 141  
Db 303 NLISKNNNN 312  
  
RESULT 7  
Q25705 ID Q25705 PRELIMINARY; PRT; 379 AA.  
AC Q25705  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE Polymorphic antigen.

```
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RL Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RT Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; U08851; AAC47831.1;
SQ SEQUENCE 379 AA; 43344 MW; DCTAF106887C8AA0 CRC64;

Query Match 14.7%; Score 110; DB 5; Length 379;
Best Local Similarity 23.4%; Pred. No. 0.86;
Matches 37; Conservative 30; Mismatches 41; Indels 50; Gaps 7;

QY 28 YGELEKGYQ-----FD---GWEISGF--EGKKDAG-----YVI 56
DB 185 YAGKEKDYERAKNAYQKQANQAVLKAKAEASSYDYLWFEFGGVPEHKKEENMLSHLYVS 244
QY 57 NLSKDTFIPVFKIEEKEEENKPTFVSKKDNPNVHNSOLN-----ESH 103
DB 245 SKDKENISKENDVLDE--KEEAETEEREELEKEEETESISEDEEEEEEKEEENE 303
QY 104 RKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141
DB 304 KKKEQKEQSNENNQKDMEA-----QNLISKNNQNN 336

RESULT 8
Q8CPK8 PRELIMINARY; PRT; 775 AA.
AC Q8CPK8;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE Penicillin-binding protein 1.
GN SE0856.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016746; AAO04453.1;
SQ Complete proteome.
KW SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;

Query Match 14.7%; Score 110; DB 16; Length 775;
Best Local Similarity 27.0%; Pred. No. 1.8;
Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;

QY 5 VTIQNGKMSSTIVSEDFILPVK-----GELEKGYQFDGW---EISGFE-----G 48
DB 625 IITIGNGKQIKQOSVKSGTKVLPHSKVLMYDSELTP--DMTGTWEDVLAFEDLTKIKVS 683
QY 49 KKDAGYVIN--ISKDTFKIPVKIEEKEEENKPTFVSKKDNPNVHNSOLN-----KKDNPNVHNSOLNES 102
DB 684 TKNGFVNTQNSISKGIK-----NKKDIEVLSAEDTDDQKTDDESDSN 730
QY 103 HKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141
DB 731 KSKKOKADEHNSNTSSTKN-----DKSNADSKNDSDD 763

RESULT 9
Q8I436 PRELIMINARY; PRT; 3008 AA.
ID Q8I436
```

```
AC Q8I436;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFE0325W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Corton C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Darden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Ruster S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929351; CAD51431.1;
KW Hypothetical protein.
SQ SEQUENCE 3008 AA; 356023 MW; 60BCBBEE15C599B4 CRC64;

Query Match 14.6%; Score 109; DB 5; Length 3008;
Best Local Similarity 32.4%; Pred. No. 9.4;
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

QY 54 YVINLSK---DTFKIPVKIEEKEEENKPTFVSKKDNPNVHNSQ---LNESHKKE 106
DB 2310 YDIELSKIEKFGASIGPVFTD--EENKEENKN--EVNKEENKEENKEENK 2366
QY 107 DLQREH---SQKSDTKDVTATVLDKNNISSK-----STNNPNK 144
DB 2367 ENKEENKEENKEENKEENKEENKEENKEENKEENKEENKEENK 2414

RESULT 10
O77355 PRELIMINARY; PRT; 600 AA.
ID O77355;
AC O77355;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Hypothetical 71.7 kDa protein.
GN PFC0465C, MAL3P4.20.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10449855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes A., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
```





```
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014834; AAN35542.1; -.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match
Best Local Similarity 14.4%; Score 108; DB 5; Length 354;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 7 IQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVNLNLS 59
DB 181 VLKAKEASS-----YDYIL-----GWFGGVPEHKKEENMLSHLYVSSKD 221

QY 60 KDTFIKPVFKTIEKKGE-----ENKPTFDVSKKDNPNQVNHSQLNSHRKE 106
DB 222 KENISKENDVDLDEKEEAETEEBELEKNEETESEISEDEEEEEEKEEENDKKK 281

QY 107 DLQREHSQKSTKDVTVATVLDKNNISKSSTNN 141
DB 282 EQEKEQSNENNDQKDMEA-----QNLISKQNNN 311

RESULT 14
QY Q815F3 PRELIMINARY; PRT; 829 AA.
AC Q815F3;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFL1275C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]_TaxID=36329;
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=2225705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Sun B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014848; AAN36341.1; -.
SQ SEQUENCE 829 AA; 98816 MW; EF2675B301B2CE93 CRC64;

Query Match
Best Local Similarity 14.3%; Score 107; DB 5; Length 829;
Matches 36; Conservative 30; Mismatches 43; Indels 36; Gaps 6;

QY 16 TVTSEEDFILPVYKGELEKGYQFDGWEISGFEGFKKDGAGVNLNLSKDTFIKPVFKTIEKK 75
DB 519 TLMTERDIVLDFHPYIMKKYHLN-----KKETLFFNLSN-----FREIEKNK 561

QY 76 EENKPT-----FDVSKKDN--PQVNHSQLNESH--RKEDLQREHSQKSDS 119
DB 562 RUKKKGTHNNKNDAAEYMLKTKKKKKNFENNNTNNKNNKLVH-----DNS 618

QY 120 TKDVTATVLDKNNISKSSTNNPNK 144
DB 619 LKQEQIINDKNVIEHTKIYDNQKK 643
```

```
RESULT 15
QY Q9FJK9 PRELIMINARY; PRT; 470 AA.
AC Q9FJK9;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Gb|AAF20218.1.
OC Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]_TaxID=3702;
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
DR EMBL; AB015468; BAB10694.1; -.
SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72E35AC54 CRC64;

Query Match
Best Local Similarity 14.2%; Score 106.5; DB 10; Length 470;
Matches 36; Conservative 31; Mismatches 65; Indels 45; Gaps 5;

QY 1 HRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF-----GKKDA 52
DB 84 NRVTDTVQNNNGSK-----YVQLARRIRYDE--EATGSQAQRIDHPNQNV 131

QY 53 GYVNLNLSKDTFIKPVFKTIEKKGEENKPTFDVSKKDN-----PQ 93
DB 132 GITEKAPENSPIETSHVEDNKRINQKNTAAKSSENAVSRVSGADHKRAEVMGKPM 191

QY 94 VNHSQLNE-----SHRKEDLQREHSQKSTKDVTVATVLDKNNISKSSTNNPNK 144
DB 192 ENRDQVRQTEAESAESHKRNVTSEKPRDQGVKKTEAKDRNKEKEEKTESINK 248

RESULT 16
QY Q9U6C4 PRELIMINARY; PRT; 379 AA.
AC Q9U6C4;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Polymorphic antigen.
GN MSP-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]_TaxID=5833;
RP SEQUENCE FROM N.A.
RC STRAIN=FC1/HN;
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;
RT "Sequence of Plasmodium falciparum secreted polymorphic antigen
RT gene.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF198190; AAF04099.1; -.
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1F9D5F25 CRC64;

Query Match
Best Local Similarity 14.0%; Score 104.5; DB 5; Length 379;
Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;

QY 7 IQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVNLNLS 59
DB 207 VLKAKEASS-----YDYIL-----GWFGGVPEHKKEENMLSHLYVSSKD 247
```

```
Qy 60 KDTFKPVFKIEKKEEENKPTFDVSKKKNPQVNHSQLN-----ESHKKE 106
Db 248 KENISKENDDDVLDE--KEEAEETEEELKEKNEEETESEI SEDEEEEBEKEEENKCK 306

Qy 107 DLQREHSQKSDSTKQVTVATVLDKNNISSKSTTN 141
Db 307 EOEKEQSNENNDQKDMEA-----ONLSKNQNNN 336

RESULT 17
Q26019 PRELIMINARY; PRT; 380 AA.
AC Q26019;
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Polymorphic antigen precursor.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RX MEDLINE=95198774; PubMed=7891748;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RA "Molecular variation in a novel polymorphic antigen associated with
RT Plasmodium falciparum merozoites.";
RT Mol. Biochem. Parasitol. 68:53-67(1994).
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN 2;
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RX MEDLINE=95198774; PubMed=7891748;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RA "Molecular variation in a novel polymorphic antigen associated with
RT Plasmodium falciparum merozoites.";
RT Mol. Biochem. Parasitol. 68:53-67(1994).
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN 2;
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RA "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RT Mol. Biochem. Parasitol. 90:21-31(1997).
RL EMBL; L07944; AAC09378.1; -.
RW Signal.
KW SIGNAL.
FT CHAIN 1 25 POTENTIAL.
FT CHAIN 26 380 POLYMORPHIC ANTIGEN.
SQ SEQUENCE 380 AA; 43290 MW; 0986CA139309CA2 CRC64;

Query Match 14.0%; Score 104.5; DB 5; Length 380;
Best Local Similarity 24.5%; Pred. No. 2.3;
Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;

Qy 7 IQNGKMSSTIVSEDFILPVYKGELEKGYQDQWEISGF--EGKQDAG-----YVINLS 59
Db 208 VLKAKEASS-----YDYIL-----GWEFGGVPEHKKEENMLSHLYVSSKD 248

Qy 60 KDTFKPVFKIEKKEEENKPTFDVSKKKNPQVNHSQLN-----ESHKKE 106
Db 249 KENISKENDDDVLDE--KEEAEETEEELKEKNEEETESEI SEDEEEEBEKEEENKCK 307

Qy 107 DLQREHSQKSDSTKQVTVATVLDKNNISSKSTTN 141
Db 308 EOEKEQSNENNDQKDMEA-----ONLSKNQNNN 337

RESULT 18
O96229 PRELIMINARY; PRT; 951 AA.
AC O96229;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF08080W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN 1;

Qy 60 KDTFKPVFKIEKKEEENKPTFDVSKKKNPQVNHSQLN-----ESHKKE 106
Db 248 KENISKENDDDVLDE--KEEAEETEEELKEKNEEETESEI SEDEEEEBEKEEENKCK 306

Qy 107 DLQREHSQKSDSTKQVTVATVLDKNNISSKSTTN 141
Db 307 EOEKEQSNENNDQKDMEA-----ONLSKNQNNN 336

RESULT 19
O44016 PRELIMINARY; PRT; 325 AA.
AC O44016;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE G5 ORF.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=NC4;
RX MEDLINE=98198836; PubMed=9539429;
RA Rieben W.K. Jr., Gonzales C.M., Gonzales S.T., Pilkington K.J.,
RA Kiyosawa H., Hughes J.E., Welker D.L.;
RA "Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to
RT the Ddp1 and Ddp2 plasmid families.";
RL Genetics 148:1117-1125(1998).
DR EMBL; U00796; AAC18634.1; -.
SQ SEQUENCE 325 AA; 38448 MW; 69A43D0C632058A6 CRC64;

Query Match 13.8%; Score 103.5; DB 5; Length 325;
Best Local Similarity 23.8%; Pred. No. 2.3;
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RA "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
RN 2;
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RA "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE001410; AAC71925.2; -.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 112486 MW; AC8D899358A84F4F CRC64;

Query Match 13.9%; Score 104; DB 5; Length 951;
Best Local Similarity 25.9%; Pred. No. 6.6;
Matches 42; Conservative 29; Mismatches 35; Indels 56; Gaps 10;

Qy 1 HRVT-VTIQNGKMSSTIVSEDFILPVYKGELEKGYQDQWEI--SGFEKKDAGYVIN 57
Db 139 HRQNELNLQSGK-----NEQDI-----NKNEKGKQ-----DISNSNAENKGD 175

Qy 58 LSKDFTFKPVFKIEKKEE-----ENKPTFD-----VSKKKNPQVNHSQLNESH 103
Db 176 -----VKEGVKELEKKEKESIDDDHKVKEENKSDDDHKVKEENKSDDDHKVKEENKSDDDH 229

Qy 104 RKEDLQR-EHSHKSDSTKQVTVATVLDKNNISSKSTTNPNK 144
Db 230 KIREVKVVEHEDEEE-----DKKKEKSENKKNKDNK 262

RESULT 19
O44016 PRELIMINARY; PRT; 325 AA.
AC O44016;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE G5 ORF.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=NC4;
RX MEDLINE=98198836; PubMed=9539429;
RA Rieben W.K. Jr., Gonzales C.M., Gonzales S.T., Pilkington K.J.,
RA Kiyosawa H., Hughes J.E., Welker D.L.;
RA "Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to
RT the Ddp1 and Ddp2 plasmid families.";
RL Genetics 148:1117-1125(1998).
DR EMBL; U00796; AAC18634.1; -.
SQ SEQUENCE 325 AA; 38448 MW; 69A43D0C632058A6 CRC64;

Query Match 13.8%; Score 103.5; DB 5; Length 325;
Best Local Similarity 23.8%; Pred. No. 2.3;
```

```
Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;
QY 1 HRTVTIQKMSSTIVSEEDFILPVYK-GEL--EKGYQDGEISGFEK-----49
DB 57 HRTITSIKN--RFSVKIGDEEKLFRISKNGELIVLNELEFDNFHFK--EGKHLRKSXMF 112
QY 50 ---KDAGVV-----INLSKDTFKPV-----FKK-----70
DB 113 NHIKSGYATNNEEIEIFLESTLCKEITAOYKRNYSKKRNIIKLPBEEBEEBEEBEE 172
QY 71 --IEBKKEENKPTFDVSKKKNQPNVHNSQLNESHKKE-----DLQREHSQKSDSTK 121
DB 173 EEEEQEEVEKPTISEEBETPAVSEBEEBEEBEEBEEBEEBEEBEEBEEBEEBEEBEE 232
QY 122 D 122
DB 233 D 233

RESULT 20
ID Q25706 PRELIMINARY; PRT; 379 AA.
AC Q25706;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Plasmodium falciparum.
DE Polymorphic antigen.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS12;
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; U08852; AAC47832.1; -.
SQ SEQUENCE 379 AA; 43302 MW; ABF9D54E1ED91A24 CRC64;

Query Match 13.8%; Score 103.5; DB 5; Length 379;
Best Local Similarity 24.5%; Pred. No. 2.7;
Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;
QY 7 IQNGKMSSTIVSEEDFILPVYKGELEKGYQDGEISGF--EGKKDAG-----YVINLS 59
DB 207 VLKAKEAS-----YDYIL-----GWFGGVPEHKKENMLSHLYVSSKD 247
QY 60 KDTFKIPVKIEEKEENKPTFDVSKKKNQPNVHNSQLN-----ESHRKE 106
DB 248 KENISKENDVLDE-KEEAEETEBELEKNEEETESISEBDEBEEBEEBEEBEEBEEBEE 306
QY 107 DLQREHSQKSDSTKDVATVLDKNNISKSSTNN 141
DB 307 EOEKESQNNNDQKDMEA-----QNLISKNNNN 336

RESULT 21
ID Q9LH98 PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=Columbia;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL; AP002057; BAB03174.1; -.
SQ SEQUENCE 2081 AA; 232851 MW; D3603E1F85EFPF29 CRC64;

Query Match 13.7%; Score 102.5; DB 10; Length 2081;
Best Local Similarity 27.4%; Pred. No. 20;
Matches 43; Conservative 24; Mismatches 63; Indels 27; Gaps 6;
QY 4 TVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDGEISGFEKGD-----AG 53
DB 1659 TVEINGGEELSTEEGSKD-----GKIEGK--EGKENSTKEGSKDDKIEEGMEGKEN 1708
QY 54 YVINLSKDTFKIPVKIEEKEENKPTFDVSK-KKNQPNVHNSQLNESHKRELQ- 109
DB 1709 STKESSKDGKINEIHGDKKATMEEGSKDGGTNGTCKDSKSKSVINGVKDDSLKDDSKN 1768
QY 110 ---REHSQKSDSTKDVATVLDKNNISKSSTNNPN 143
DB 1769 GDINEINGKEDSVKDNVTETIQGNDNSLTNSTSSEP 1805

RESULT 22
ID Q9V7J0 PRELIMINARY; PRT; 382 AA.
AC Q9V7J0; Q9Q081;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CG8421 protein (Aspartyl beta-hydroxylase variant 2).
OS ASPH OR CG8421 OR CG18658.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
```

```
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Fieriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mirza S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smurniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RA "Aspartyl beta -Hydroxylase (Asph) and an Evolutionarily Conserved
RT Isoform of Asph Missing the Catalytic Domain Share Exons with
RT Junction.";
RL J. Biol. Chem. 275:39543-39554(2000).
DR EMBL; AF289494; AAF58063.2; -
DR EMBL; AF289494; AAG40807.1; -
DR FlyBase; FBgn0034075; Asph.
SQ SEQUENCE 382 AA; 43287 MW; 60B5C03ABBF6C6E8B CRC64;
Query Match 13.6%; Score 101.5; DB 5; Length 382;
Best Local Similarity 24.5%; Pred. No. 3.8;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;
OY 21 EDFILPVYKLEKGYQFDGW-----EISFGKKDAGYVI-----NLKDTFIK 65
DB 78 EDLDTPLSESFSK--VFDGWVDEHRDGHVQPSGEALDDHDEHDDHDEDEDE 135
```

J. Biol. Chem. 275:39543-39554 (2000).

[2] SEQUENCE FROM N.A.

RP STRAIN=Berkelley;

RC MEDLINE=20196066; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Botchan P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Flossler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacלב J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spiers E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*."

RL Science 287:2185-2195 (2000).

[3]

RP SEQUENCE FROM N.A.

RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacלב J., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RT "Sequencing of *Drosophila melanogaster* genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[4]

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradscky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,

RA Tupy J.B., Bergman C., Berman B., Carlson J.W., Celniker S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

RT "Annotation of *Drosophila melanogaster* genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[5]

RP SEQUENCE FROM N.A.

RX MEDLINE=22255705; PubMed=12368864;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan J.-S., Nene V., Shallow J.J., Suh B., Peterson J., Angiuoli S.,  
RA Perera M., Allen J., Sellenent J., Haft D., Macher M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.R.W.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrrell B.,  
RA "Genome sequence of the human malaria parasite *Plasmodium*  
RT *falciparum*.";





RA Perriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattel B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Shapleton M., Strong R., Svazkas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RL "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berlan B., Carlson J.W., Celniker S.E.,  
 RA Klomp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Crumiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RL "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003578; AAP51062.2; -  
 DR HSP; P02593; ICTR.  
 DR FlyBase; FBgn0051958; CG31958.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF000036; ehand; 3.  
 DR ProDom; PD000012; EF-hand; 1.  
 DR SMART; SM00054; Eph; 3.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 KW Calcium-binding.  
 SQ SEQUENCE 157 AA; 18263 MW; 543F0480E11D9EC1 CRC64;  
  
 Query Match 13.1%; Score 98; DB 5; Length 157;  
 Best Local Similarity 22.7%; Pred.No. 2.7;  
 Matches 34; Conservative 29; Mismatches 53; Indels 34; Gaps 5;  
  
 Qy 13 MSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKQAGYVI-----NLSKDTF 63  
 Db 1 MDELSVEEQDLKNYI-SLLDK-----DNEGAITSKEIGWIRALGRQPNESIAKEEF 52  
  
 Qy 64 IKPVFKKIEE-KGEENKPTFDVSKKDN-----PQVNHISQLNESHKE 106  
 Db 53 CNVILRKMDHNTKEEELDAFRVFDKENNGYISTELRAVFMALGKLEDELEEMIREY 112  
  
 Qy 107 DLQREEHSQKSDTKDVTATVLKNNISSK 136  
 Db 113 DLQDDHNFEEENNMMWTTPIILNHNHTAVK 142  
  
 RESULT 32  
 Q8IHW3 PRELIMINARY; PRT; 449 AA.  
 ID Q8IHW3  
 AC Q8IHW3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PF11\_0413.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 ON NCBI\_TaxID=36329;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX

Db 194 EGTVEATVEATTEATTAT 212

RESULT 34

Q815T1 PRELIMINARY; PRT; 558 AA.

AC Q815T1;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein.

GN PFJ0600W.

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=36329;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=3D7;

RX MEDLINE=22255705; PubMed=12369864;

RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,

RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,

RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,

RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,

RA Pertea M., Allen J., Selengut J., Hart D., Mather M.W., Vaidya A.B.,

RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Vaidya A.B.,

RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,

RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

RA Fraser C.M., Barrell B.

RT "Genome sequence of the human malaria parasite Plasmodium falciparum.";

RL Nature 419:498-511(2002).

DR EMBL; AE014846; AAN36209.1; -.

KW Hypothetical protein.

SQ SEQUENCE 558 AA; 66487 MW; 2ADC29EA69FF90D3 CRC64;

Query Match 13.0%; Score 97; DB 5; Length 558;

Best Local Similarity 22.4%; Pred. No. 13;

Matches 35; Conservative 31; Mismatches 64; Indels 26; Gaps 4

QY 8 QNGKEMSTVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVNLNKDTFIKPV 67

Db 41 QNGKKYIYNLSKQNNKIIGGSKHNGHLF-----INNYKKRNKYKINKYKPCISIFS 96

QY 68 FKITEEKKEENKPTFDVSKKDPQVNHSQLNE-----SHRKEDL 108

Db 97 FPK-KDKNKNDKN--DKTHMKDNSLENKEKEIGKDKQEKKEKIILKDNKKLDE 153

QY 109 QREBHSQSDSKDVTATVLDKNNISSKSTNNPNK 144

Db 154 KKKELQKEKKTKLTIVLQKNNKNNNNNNNNK 189

RESULT 35

Q9STN4 PRELIMINARY; PRT; 988 AA.

AC Q9STN4;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Hypothetical 109.0 kDa protein.

GN T28D5.30 OR AT4G08340.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,

RA Barrell B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K.,

RA Schueller C.;

RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

```

RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL109819; CAB52556.1; --
DR EMBL; AL161511; CAB77959.1; --
DR InterPro; IPR001760; Opsin.
DR InterPro; IPR003653; SUMO protease.
DR Pfam; PF02902; Peptidase_C48; 1.
DR PROSITE; PS00238; OPSIN; 1.
DR PROSITE; PS06600; ULP_PROTEASE; 1.
KW Hypothetical protein.
SQ SEQUENCE 988 AA; 108955 MW; 01C518587D460EAD CRC64;

Query Match 13.0%; Score 97; DB 10; Length 988;
Best Local Similarity 22.1%; Pred. No. 24;
Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;

Qy 9 NGKMSSTIVSEEDFILPVYKGELEKGY-----QFDGWEISGFEGKDGAGVYINLSKDTF 63
Db NGRQNSNVQSVDEILSYTDKVGGLNVSERDVELDEDDVRSAGGLSPVQRDN- 441
Qy 64 IKPVFKKIEEKE-----EENKPTFDVSKKDNQVNHSQLNSHREKDLQREE 112
Db VEPVGDVDRSGSDMSPNSAANNVREGPATFDIMESDNPGRDNVAPWEDHIRSRVQLSP 501
Qy 113 HSQKSDSTKDYT--ATVLDKNNISSKSTTNNPNK 144
Db HVL-----GAKDVTDSVSDPTDKVGVDVTDASDPTE 532

RESULT 36
Q81226 PRELIMINARY; PRT; 238 AA.
AC Q81226;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF10765W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929351; CAD51442.1; --
KW Hypothetical protein.
SQ SEQUENCE 531 AA; 62575 MW; 3A77BD86BEBE4A41 CRC64;

Query Match 12.8%; Score 96; DB 5; Length 531;
Best Local Similarity 27.7%; Pred. No. 14;
Matches 44; Conservative 27; Mismatches 64; Indels 24; Gaps 8;

Qy 5 VTIONGKMSSTIVSEEDFI-LPVYKGELEKGYFDGWEI-----SGFEGKDGAGVYI 56
Db IEIENKQLIDLKKKIBELLSVPLEKQEL-----YLLDSPQILLKDNVYKLRCKIQNGSMI 73
Qy 57 NLSKDTFTKPVFKIE--EKKEEN-KPTFDVSKKDNQVNHSQLNE--SHRKEDL--- 108
Db 74 QLKSD--IKPSTKKKEENKKNENIKNTDENNKETYDKDVHKNQMDSLCQKKEIVNN 131
Qy 109 ---QREHSQKSDSTKDYTATVLDKNNISSKSTTNNPNK 144

```

```

SQ SEQUENCE 238 AA; 28481 MW; DA175A2B5A109B49 CRC64;

Query Match 12.8%; Score 96; DB 5; Length 238;
Best Local Similarity 28.7%; Pred. No. 6.1;
Matches 45; Conservative 24; Mismatches 58; Indels 30; Gaps 9;

Qy 8 QNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISG---FEGKDA-----GYVIN 57
Db 51 ENKSNVNVSSSENF-MKIYKNILKQNYHVEFKQKNENIEESKKKATSKKTKLNFK 109
Qy 58 LSKDTFTKPVFKIEEKEENKPTFDVSKKDN--QV--NH--SOLNESHREKED--- 107
Db 110 LKKNLLEKVIK---KAFLEAKNTVPFSLKYYTVPHISNNHNISQNDHKKKQDKNIK 166
Qy 108 -LQREHSQKSDSTKDYTATVLDKNNISSKSTTNNPN 143
Db 167 LLKLEKSNKDDNN-----NINNKNNKQKCDTKNLN 198

RESULT 37
Q81426 PRELIMINARY; PRT; 531 AA.
AC Q81426;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein, conserved.
GN PF0380C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=2255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929351; CAD51442.1; --
KW Hypothetical protein.
SQ SEQUENCE 531 AA; 62575 MW; 3A77BD86BEBE4A41 CRC64;

Query Match 12.8%; Score 96; DB 5; Length 531;
Best Local Similarity 27.7%; Pred. No. 14;
Matches 44; Conservative 27; Mismatches 64; Indels 24; Gaps 8;

Qy 5 VTIONGKMSSTIVSEEDFI-LPVYKGELEKGYFDGWEI-----SGFEGKDGAGVYI 56
Db 17 IEIENKQLIDLKKKIBELLSVPLEKQEL-----YLLDSPQILLKDNVYKLRCKIQNGSMI 73
Qy 57 NLSKDTFTKPVFKIE--EKKEEN-KPTFDVSKKDNQVNHSQLNE--SHRKEDL--- 108
Db 74 QLKSD--IKPSTKKKEENKKNENIKNTDENNKETYDKDVHKNQMDSLCQKKEIVNN 131
Qy 109 ---QREHSQKSDSTKDYTATVLDKNNISSKSTTNNPNK 144

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 30.4847 Seconds  
(without alignments)  
697.707 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSSITVSEDFILPVYK.....ATVLDKNISKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	696	100.0	773	22 AAB48343	S. pneumoniae Sp13
2	696	100.0	2120	21 AAY81710	Streptococcus pneu
3	696	100.0	2140	24 ABU01020	S. pneumoniae type
4	615	88.4	117	19 AAW55096	Streptococcus pneu
5	615	88.4	117	23 ABP54590	S. pneumoniae Sp04
6	101.5	14.6	564	22 ABB61977	Drosophila melanog
7	98	14.1	665	21 AAB18278	Plasmodium falcipa
8	92.5	13.3	2519	22 ABG15636	Novel human diagno
9	91	13.1	484	21 AAG47777	Arabidopsis thalia

10	89.5	12.9	511	20	AAV35091	Chlamydia pneumonia
11	88.5	12.7	1141	22	AAG85008	Shrimp white spot
12	88	12.6	225	23	ABP73992	Candida albicans e
13	88	12.6	258	22	AAB94584	Human protein sequ
14	88	12.6	746	22	AAG81779	S. epidermidis ope
15	88	12.6	778	23	ABF39023	Staphylococcus epi
16	88	12.6	817	22	AAW79318	Human protein SEQ
17	88	12.6	817	22	AAW79319	Human protein SEQ
18	87	12.5	281	22	ABB64828	Drosophila melanog
19	87	12.5	607	22	AAV39950	Human polypeptide
20	87	12.5	635	21	AA18176	Plasmodium falcipa
21	86.5	12.4	2485	21	AA18176	Plasmodium falcipa
22	86	12.4	1980	21	AA18176	Human sodium chann
23	86	12.4	1980	23	AAQ14927	Human sodium chann
24	85.5	12.3	884	23	ABP73375	Candida albicans e
25	85	12.2	329	21	AAG08296	Arabidopsis thalia
26	85	12.2	339	21	AAG08295	Arabidopsis thalia
27	85	12.2	358	21	AAG08294	Arabidopsis thalia
28	85	12.2	465	23	ABF39011	Staphylococcus epi
29	85	12.2	472	22	AAG81782	S. epidermidis ope
30	85	12.2	720	22	AAG82482	S. epidermidis ope
31	85	12.2	728	23	ABP40123	Staphylococcus epi
32	85	12.2	1166	21	AA18288	Plasmodium falcipa
33	85	12.2	1978	19	AAW69361	Tetradotoxin-sensi
34	85	12.2	1988	19	AAW69362	Tetradotoxin-sensi
35	84.5	12.1	157	23	ABP29728	Streptococcus poly
36	84.5	12.1	161	23	ABP26418	Streptococcus poly
37	84.5	12.1	309	21	AAG35428	Zea mays protein f
38	84.5	12.1	361	21	AAG35427	Zea mays protein f
39	84.5	12.1	390	21	AA183503	Protein involved i
40	84.5	12.1	416	21	AAG35426	Zea mays protein f
41	84.5	12.1	434	23	ABP73309	Candida albicans e
42	84.5	12.1	645	24	ABU19106	Pathogen specific
43	84.5	12.1	654	24	ABP56879	Staphylococcus epi
44	84.5	12.1	2663	22	ABG98612	Human tumour suppr
45	84	12.1	215	23	ABU11001	Yeast selected int

#### ALIGNMENTS

##### RESULT 1

AAB48343  
ID AAB48343 standard; Protein; 773 AA.

XX AAB48343;

AC AAB48343;

XX 20-APR-2001 (first entry)

XX S. pneumoniae Sp130 polypeptide.

XX Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;

XX bronchial; lung; blood; infection; immune response; immunotherapy;

XX antibacterial; auditory; vaccine.

XX Streptococcus pneumoniae.

XX WO2000076540-A2.

XX 21-DEC-2000.

XX 09-JUN-2000; 2000WO-US15925.

XX 10-JUN-1999; 99US-0138453.

XX (MEDI-) MED IMMUNE INC.

XX Adamou JE, Choi GH;

XX WPI; 2001-112197/12.

XX N-PSDB; AAC84742.

XX New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by  
 PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
 XX blood infections  
 XX  
 PS Claim 8; Page 51-54; 54pp; English.  
 CC The invention relates to novel immunogenic polypeptides, Sp128 and Sp130  
 CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
 CC the treatment and prevention of pneumococcal infections, particularly  
 CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
 CC bronchial, lung or blood infections. The antigens are used as immunogenic  
 CC agents to stimulate an immune response. The antisera and antibodies may  
 CC also be used in diagnosing and treating pneumococcal infections.  
 CC Recombinant polypeptides serve as a mechanism for stimulating production  
 CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
 CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the S. pneumoniae Sp130 polypeptide.  
 XX  
 SQ Sequence 773 AA;

Query Match 100.0%; Score 696; DB 22; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-63;  
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 60  
 DB 640 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 699  
 QY 61 IEKKEEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 120  
 DB 700 IEKKEEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 759  
 QY 121 NNISKSSTNNPNK 134  
 DB 760 NNISKSSTNNPNK 773

RESULT 2  
 ID AAY81710  
 AC AAY81710 standard; Protein; 2120 AA.  
 AC AAY81710;  
 XX  
 DT 02-JUN-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae protein sequence ID3.  
 XX  
 KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200006738-A2.  
 XX  
 PD 10-FEB-2000.  
 XX  
 XX 27-JUL-1999; 99WO-GB02452.  
 PF  
 XX 27-JUL-1998; 98GB-0016336.  
 PR  
 XX 19-MAR-1999; 99US-0125329.  
 PR  
 XX (MICR-) MICROBIAL TECHNIQS LTD.  
 PA  
 XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
 PI  
 XX WPI; 2000-195301/17.  
 DR  
 XX N-PSDB; AA291806.  
 DR  
 XX Streptococcal proteins and polynucleotides useful for diagnosis,  
 PT treatment and prophylaxis of bacterial infections  
 PT  
 XX

PS Claim 2; Page 41-42; 76pp; English.  
 XX  
 CC This sequence represents a Streptococcus pneumoniae protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful as vaccines and also in  
 CC diagnostic assays. The sequences are useful for the detection or  
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
 CC infection. As the sequences can be used to treat S. pneumoniae infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
 CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
 CC meningitis.  
 XX  
 SQ Sequence 2120 AA;

Query Match 100.0%; Score 696; DB 21; Length 2120;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-62;  
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 60  
 DB 1953 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 2012  
 QY 61 IEKKEEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 120  
 DB 2013 IEKKEEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 2072  
 QY 121 NNISKSSTNNPNK 134  
 DB 2073 NNISKSSTNNPNK 2086

RESULT 3  
 ID ABU01020  
 AC ABU01020 standard; Protein; 2140 AA.  
 AC ABU01020;  
 XX  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain protein from coding region #590.  
 XX  
 KW Bacterial meningitis; pneumonia; sepsis; otitis media;  
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KW auditory; respiratory; gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae type 4 strain.  
 XX  
 PN WO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 XX 27-MAR-2002; 2002WO-IB02163.  
 PF  
 XX 27-MAR-2001; 2001GB-0007658.  
 PR  
 XX (CHIR-) CHIRON SPA.  
 PA  
 XX (GENO-) INST GENOMIC RES.  
 PA  
 XX Masignani V, Tettelin H, Fraser C;  
 PI  
 XX WPI; 2003-040579/03.  
 DR  
 XX N-PSDB; ABX06302.  
 DR  
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
 PT

```
PT or ear infection -
XX Claim 1; SEQ ID No 1180; 56pp; English.
XX
XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 of 2489 identified DNA coding regions from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX ABS56454. Also included are an antibody which binds one of the
XX proteins, treating a patient by administering the protein, DNA or
XX antibody (in a composition), a kit comprising first and second primers,
XX which are the nucleic acid cited above or fragments between nucleotides
XX 8-100 of a sequence not defined in the specification, for amplifying a
XX target sequence contained within a Streptococcus nucleic acid sequence,
XX where the first primer is substantially complementary to the target
XX sequence and the second primer is substantially complementary to the
XX complement of the target sequence, and where the parts of the primers
XX having substantial complementarity define the termini of the target
XX sequence to be amplified, assay comprising contacting a test compound
XX with the protein, and determining whether the test compound binds to the
XX protein and a Streptococcus pneumoniae bacterium, where one or more
XX genes encoding the proteins has been rendered inactive. The proteins,
XX nucleic acid molecules, antibody and compositions are useful as
XX medicaments for treating or preventing a disease or infection due to
XX streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
XX sepsis, otitis media or ear infection. They are also useful in developing
XX vaccines, diagnostics and antibiotics. The methods are useful for
XX identifying immunodominant proteins. The present sequence is one of
XX the 2469 proteins expressed by the identified coding regions from the
XX genomic sequence.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2140 AA;
Query Match 100.0%; Score 696; DB 24; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.3e-62; Mismatches 0; Indels 0; Gaps 0;
Matches 134; Conservative 0;
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKK 60
DB 1973 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKK 2032
QY 61 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120
DB 2033 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 2092
QY 121 NNISKSSTNNPNK 134
DB 2093 NNISKSSTNNPNK 2106
RESULT 4
AAW55096
ID AAW55096 standard; Protein; 117 AA.
XX
XX AAW55096;
XX
XX 02-OCT-1998 (first entry)
XX Streptococcus pneumoniae SP0043 protein.
XX
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX detection; pneumonia; otitis media; meningitis.
XX
XX Streptococcus pneumoniae.
XX
XX WO9818930-A2.
XX
XX 07-MAY-1998.
XX
```

---

```
XX 30-OCT-1997; 97WO-US19422.
XX
XX 31-OCT-1996; 96US-0029960.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX WPI; 1998-272224/24.
XX N-PSDB; RAV27357.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
XX pneumoniae - or their epitope-containing fragments, useful in
XX protective or therapeutic vaccines, and for diagnosis
XX
XX Claim 11; Page 62; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX can be useful in vaccines for inducing protective antibodies against
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX are used to detect Streptococcus infection (by usual hybridisation or
XX amplification methods), also for isolating Streptococcus genes or their
XX allelic variants. The protein can be used similarly to detect specific
XX antibodies in standard immunoassays, especially for diagnosing or
XX monitoring infections. Antibodies which bind the protein are used to
XX detect corresponding antigens, to purify the protein and for passive
XX immunisation (optionally coupled to a toxin). Vaccines are administered,
XX e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose.
XX
XX Sequence 117 AA;
Query Match 88.4%; Score 615; DB 19; Length 117;
Best Local Similarity 100.0%; Pred. No. 6.1e-56;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 YKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 77
DB 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 60
QY 78 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSSTNNPNK 134
DB 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSSTNNPNK 117
RESULT 5
*ABP54590
ID ABP54590 standard; Protein; 117 AA.
XX
XX ABP54590;
XX
XX 04-SEP-2002 (first entry)
XX
XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.
XX
XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX antibacterial; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
XX
XX US2002061545-A1.
XX
XX 23-MAY-2002.
XX
XX 22-JAN-2001; 2001US-0765272.
XX
XX 30-OCT-1997; 97US-0961083.
XX
XX (CHOI/) CHOI G H.
XX (KUNS/) KUNS C A.
XX
```







PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142877.  
PR 13-JUL-1999; 99US-0143342.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

```

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160983.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161362.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 13.1%; Score 91; DB 21; Length 484;
Best Local Similarity 19.3%; Pred. No. 0.89;
Matches 29; Conservative 29; Mismatches 59; Indels 34; Gaps 4;

Qy 18 YKGELEKGYQDFWEISGFE-----GKIDAGVIVNLKDTIKPVFKKIEKKKEEN 69
Db 114 YVQDLARRIRYDE-EATGSAQRIDHPNKNVIGITEKAFENSPTEETSHRVDDNKRINN 172

Qy 70 KPTFDVSKKDN-----PQVNHSLN-----SHRKEDLQREHS 104
Db 173 QKNFTAAKSSNAVSRVSGADHKRAEVMGKPMENRDQVOTESAEKSHRKNVTKSEK 232

Qy 105 QKSDTKDVTATVLDKNNISSKSTNNPNK 134
Db 233 RDQGVKYTEAKDKDRNKEKKEKTESINK 262

RESULT 10
AAV35091
ID AAV35091 standard; Protein; 511 AA.
XX
AC AAV35091;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae transmembrane protein sequence.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST ) GENSET.

```

```

XX Griffais R;
XX PI
XX DR WPI; 1999-357842/30.
XX PT
XX PS Genome sequence of Chlamydia pneumoniae
XX PS Page 975-976; Disclosure; 1912pp; English.
XX CC AAY34584-Y35879 represent the proteins encoded by all the open reading
XX CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
XX CC C. pneumoniae causes respiratory disease such as pneumonia and
XX CC bronchitis and is thought to be a contributing factor in heart
XX CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
XX CC nodosum or pharyngitis. The polypeptides encoded by the open reading
XX CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
XX CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
XX CC nucleotide sequences can also be used as immunogenic compositions,
XX CC especially where the vector directs the expression of a neutralising
XX CC epitope of C. pneumoniae.
XX SQ Sequence 511 AA;

Query Match 12.9%; Score 89.5; DB 20; Length 511;
Best Local Similarity 24.5%; Pred. No. 1.4;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

Qy 54 IKPVFKKIEKKKEENKPTFD-----VSKKDNQVNHSLNESHK 95
Db 95 VKGVFKTTPQARPEVSSPRLPSHVQHGQRLPGLGFRDQKRSNEDADLGKMKRSYSD 154

Qy 96 EDLQREHSQKSDTKDVTATVLDKNNISSKSTT 129
Db 155 GDLDRVGHDSNEDSTEDSR-----EGGEPSKSSS 185

RESULT 11
AAG85008
ID AAG85008 standard; Protein; 1141 AA.
XX
AC AAG85008;
XX
DT 11-SEP-2001 (first entry)
XX
DE Shrimp white spot Bacilliform virus (WSBV) protein 99.
XX
KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW antiviral agent; gene expression; antisense construct;
KW transgenic viral resistant shrimp.
XX
OS White spot syndrome virus.
XX
PN WO200138351-A2.
XX
PD 31-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US28888.
XX
PR 24-NOV-1999; 99CN-0124717.
XX
PA (PENY-) PE CORP NY.
XX
PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
XX
PA (SINO-) SINOGENOMAX CO LTD.
XX
PI Xu X, Yang F, He J, Pham L, He M., Ye Y, Shen Y, Kodira C;
XX WPI; 2001-355877/37.
XX DR N-PSDB; AAH62788.
XX
PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT (WSBV), useful for producing viral polypeptides that can be used to
PT screen for agents that are useful for treating WSBV infection -
XX

```





```

PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2002-381255/41.
DR N-PSDB; ABN91568.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 3868; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
XX Sequence 778 AA;
SQ
Query Match 12.6%; Score 88; DB 23; Length 778;
Best Local Similarity 25.7%; Pred. NO. 3.5;
Matches 37; Conservative 21; Mismatches 50; Indels 36; Gaps 7;
QY 4 SSTIIEEDFILPVYKGELEKGYQPDGW---EISGPE-----GKKDAGYVIN--LSKD 51
DB 643 SGTKVLPHSKVLMITDGLTMP-DMTGWTKEVDLAFEDLTJKVSTKGNQFVNTQSISKG 701
QY 52 TPIKPVFKKIEBKKEENKPTFDVS-----KKDNPQVNHSQLNESHKEDLQREHSQKS 107
DB 702 QIHK-----NKDKIEVLSLAEDTDDQEKTEDESSDNKSKKKADEHDSNTS 748
QY 108 DSTKVATVLDKNNISSKSTNN 131
DB 749 SSTKN-----DKSNADSKNDSDD 766
RESULT 16
AAM79318
ID AAM79318 standard; Protein; 817 AA.
XX
AC AAM79318;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 2964.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
DE WO200157190-A2.
XX
PN WO200157190-A2.
XX
XX 09-AUG-2001.
PD
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK52451.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 215; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 817 AA;
SQ
Query Match 12.6%; Score 88; DB 22; Length 817;
Best Local Similarity 27.3%; Pred. NO. 3.7;
Matches 33; Conservative 21; Mismatches 55; Indels 12; Gaps 4;
QY 20 GELEKGYQFDGWEISG--FEKGDAGYVINLSKDTFIRPVFKKIEKKEENKPTFDVSK 77
DB 196 GOEKQESFKSWEASGKHQESKPAVLEQRKQDTSKLRSTLPEQKKQEKSKSPSPSQ 255
QY 78 -KKDNPQVNHSQLNESHKED-----LQREHSQKSDSTKVATVLDKNNISSKST 128
DB 256 WKQDTPKSKAGYVQEEHKKQETPKLPVQLOKEQ-DPKKQTPKSWTPSMQSQNTTKSWT 314
QY 129 T 129
DB 315 T 315
RESULT 17
AAM79319
ID AAM79319 standard; Protein; 817 AA.
XX
AC AAM79319;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 2965.
XX
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
DE WO200157190-A2.
XX
PN WO200157190-A2.
XX
XX 09-AUG-2001.
PD
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.

```



```
PF 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB; AAI59106.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 3095; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAW42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 607 AA;
SQ
Query Match 12.5%; Score 87; DB 22; Length 607;
Best Local Similarity 33.3%; Pred. No. 3.2;
Matches 20; Conservative 12; Mismatches 28; Indels 0; Gaps 0;
QY 49 SKDTFTKPVFKTEKEEENKPTFDVSKKQNPQVNHSQLNESHKEDLQREHSQKSD 108
DB 373 SPDNCCNELFKKKKKKEEKKKEEKKRRRRRRKREKREKREKREKREKRD 432
RESULT 20
AAB18176
ID AAB18176 standard; Protein; 635 AA.
XX
XX AAB18176;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:33.
DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
XX Plasmodium falciparum.
OS
XX WO200025728-A2.
PN
XX 11-MAY-2000.
PD
XX 05-NOV-1999; 99WO-US26796.
PF
```

```
XX 05-NOV-1998; 98US-0107131.
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI WPI: 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 85-86; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAT70078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 635 AA;
SQ
Query Match 12.5%; Score 87; DB 21; Length 635;
Best Local Similarity 23.8%; Pred. No. 3.4;
Matches 30; Conservative 20; Mismatches 52; Indels 24; Gaps 4;
QY 22 LEKGYQFDGWAISGEGKKDAGYVNLKSKDTFKFVKFKBEKEEENKPTFDVSKKDN 81
DB 202 LQKKNYIQQDEEDNETIRSDSKLDIYSDSQSKDMMSSSPNKEES-----MSSDNHN 256
QY 82 PQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLD-----KNNISKS 127
DB 257 KDINSS---DNQNKDINSDDHM-NDSTNESTTSLSTSLSTNTNRNKKRKNINNN 311
QY 128 TTNPN 133
DB 312 NNNNSN 317
RESULT 21
AAB18172
ID AAB18172 standard; Protein; 2485 AA.
XX
XX AAB18172;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:29.
DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
XX Plasmodium falciparum.
OS
```



XX WO200025728-A2.  
XX 11-MAY-2000.  
XX 05-NOV-1999; 99WO-US26796.  
XX 05-NOV-1998; 98US-0107131.  
XX (HOFF/) HOFFMAN S.  
XX (CARU/) CARUCCI D.  
XX (GARD/) GARDNER M.  
XX (VENT/) VENTER J C.  
XX Hoffman S, Carucci D, Gardner M, Venter JC;  
XX WPI; 2000-365347/31.  
XX Proteins encoded by chromosome 2 of the human malarial parasite,  
XX Plasmodium falciparum, useful as antimalarial vaccines and in the  
XX diagnosis of P.falciparum infection -  
XX Disclosure; Page 75-82; 577pp; English.  
XX The present invention describes proteins and their fragments (I) encoded  
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
XX Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
XX vaccines against P. falciparum infection comprising (I) or (II).  
XX (I) and (II) are useful for the development of vaccines against  
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
XX antibody raised to immunogens comprising the sequences of (I), are  
XX useful in the detection of infection with P. falciparum. Furthermore,  
XX (I) (especially when they are rifins or secreted or membrane proteins)  
XX can aid the identification of drugs to treat or prevent P. falciparum  
XX infection, or they can be used to identify drug resistance in  
XX P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
XX subsequent identification of proteins encoded by it will help to expand  
XX our understanding of parasite biology, a process hampered by the  
XX complexity of the parasitic lifecycle, and provide new targets for  
XX vaccine and drug development. Parasite resistance to drugs and mosquito  
XX resistance to insecticides have led to a resurgence of malaria in many  
XX parts of the world, and there is a pressing need for vaccines and new  
XX drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
XX and protein sequences given in the present invention, but which are not  
XX specifically mentioned within the specification.  
XX Sequence 2485 AA;  
Query Match 12.4%; Score 86.5; DB 21; Length 2485;  
Best Local Similarity 23.8%; Pred. No. 24;  
Matches 36; Conservative 27; Mismatches 31; Indels 57; Gaps 9;  
Qy 21 ELEKGYQPDG-----EISGFE-----GKDAGYVINLSKDTFTKPVFKIEE 63  
Db 1977 ENERDEKDGVEEMNGDKNEEMNGDKNEEMNGDKNGG--IN-----EE 2020  
Qy 64 KK-----EENKPTFVSKKONPQV-----HSQLNESHKEDLQ-----RSEH---- 103  
Db 2021 HKNEGINEEHKDEL-INKEHKNERINEEHKNERINEEHKNEGINEEHKNER 2079  
Qy 104 ---SOKSDTKDVTATVLDKNNISKSTTN 131  
Db 2080 INEHNKNGINKLTHNNKNNISNNYND 2110  
RESULT 22  
ID AAB23563  
XX AAB23563 standard; Protein; 1980 AA.  
XX AC AAB23563;  
XX DT 09-JAN-2001 (first entry)  
XX

DE Human sodium channel 8A (SCN8A) protein sequence.  
XX Sodium channel 8A; SCN8A; human; excitatory cell.  
XX Homo sapiens.  
XX JP2000201684-A.  
XX 25-JUL-2000.  
XX 11-JAN-1999; 99JP-0004645.  
XX 11-JAN-1999; 99JP-0004645.  
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
XX WPI; 2000-545978/50.  
XX N-PSDB; AAA93791.  
XX New sodium channel SCN8A useful for the elucidation physiological  
XX mechanism participated by excitatory cells -  
XX Disclosure; Page 5-12; 14pp; Japanese.  
XX This sequence represents the human sodium channel SCN8A protein. The  
XX invention also provides the SCN8A encoding gene sequence and an antibody  
XX against the sodium channel. The SCN8A protein alpha subunit and its  
XX encoding cDNA can be used in the elucidation of the physiological  
XX mechanisms of excitatory cells.  
XX Sequence 1980 AA;  
Query Match 12.4%; Score 86; DB 21; Length 1980;  
Best Local Similarity 26.2%; Pred. No. 20;  
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;  
Qy 27 QFDGWEISGFGKKGAGYVINLSKDTFTKPVFKIEEKKENKPTFVSKKONPQVNH 86  
Db 993 EMNNLQISVIRIKKGVAWT-KLVHAFMQAHFK---QREADEVKPLDELYEKKANCIANH 1048  
Qy 87 SOLNESHKEDLQREHSQKSDSTKQVTATVLDKNNISKSTTNPN 133  
Db 1049 TGA-DIHRNGDFQKNGTGTSGIGSSVEKVIIDEDHM---SFINPN 1091  
RESULT 23  
AAO14927  
ID AAO14927 standard; Protein; 1980 AA.  
XX AAO14927;  
XX 19-JUL-2002 (first entry)  
XX Human sodium channel subunit SCN8A.  
XX Human; sodium channel subunit; SCN12A; SCN12A-s; SCN8A;  
XX human nervous system; chromosome 12q13.1; excitatory cell;  
XX drug development; familial hyperglycaemia; QT extending syndrome type 3;  
XX motor endplate disease.  
XX Homo sapiens.  
XX WO200190355-A1.  
XX 29-NOV-2001.  
XX 11-JUL-2000; 2000WO-JP04629.  
XX 23-MAY-2000; 2000JP-0152085.  
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX Kanazawa I, Goto J, Jeong S;  
PI

XX WPI; 2002-393394/42.  
DR N-PSDB; AAL42751.  
XX Sodium channel alpha subunits SCN12A and SCN8A of human nervous system  
PT together with splicing variants, useful in studying physiological  
PT mechanism relating to excitatory cells and in drug development -  
XX  
XX Claim 3; Page 97-110; 118pp; Japanese.  
XX The invention comprises the amino acid and coding sequence of three  
CC sodium channel subunits from the human nervous system (SCN12A, SCN12A-s  
CC and SCN8A). The SCN12A/SCN12A-s gene is found on chromosome 3p23-21.3 and  
CC the SCN8A gene is found on chromosome 12q13.1. The sodium ion channel  
CC subunit proteins are useful in studying physiological mechanisms relating  
CC to excitatory cells and in drug development. The sodium ion channel  
CC subunit proteins are useful for treating diseases such as familial  
CC hyperglycaemia, QT extending syndrome type 3 and motor endplate diseases.  
CC The present amino acid sequence represents the human SCN8A sodium channel  
CC subunit.  
XX  
XX Sequence 1980 AA;  
Query Match 12.4%; Score 86; DB 23; Length 1980;  
Best Local Similarity 26.2%; Pred. No. 20;  
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;  
QY 27 QDGEISGEGKQAGYVNLKDTFKVPFKIEEKEENKPTFDVSKKKNPQVNH 86  
DB 993 EMNNLQISVIRIKKGVAWT-KLVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1048  
QY 87 SOLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPN 133  
DB 1049 TGA-DIHRNGDFQKNGGTGSGVSKYIIDEHDM---SPINPN 1091  
RESULT 24  
ABP73375  
ID ABP73375 standard; Protein; 884 AA.  
XX  
AC ABP73375;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Candida albicans essential protein SEQ ID NO 7212.  
XX  
KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;  
KW signal transduction; DNA replication; cell division; growth;  
KW proliferation; Candida albicans; fungicide; antifungal.  
XX  
OS Candida albicans.  
XX  
PN WO200253728-A2.  
XX  
PD 11-JUL-2002.  
XX  
XX 26-DEC-2001; 2001WO-US9486.  
XX  
PR 29-DEC-2000; 2000US-259128P.  
PR 20-FEB-2001; 2001US-0792024.  
PR 22-AUG-2001; 2001US-314050P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
XX  
XX WPI; 2002-566694/60.  
DR N-PSDB; ABZ31925.  
XX  
XX Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele  
PT of a gene and placing other allele of the gene under conditional  
PT expression -

XX Claim 44; SEQ ID NO 7212; 167pp + Sequence Listing; English.  
PS The invention relates to constructing (M1) a strain of diploid fungal  
XX cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination, of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal  
CC cells having both alleles modified are useful for identifying a gene that  
CC is essential to the survival or growth of a fungus, a gene that  
CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
CC that contributes to the resistance of a diploid fungus to an antifungal  
CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthetic, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
XX the European Patent Office.  
XX  
XX Sequence 884 AA;  
Query Match 12.3%; Score 85.5; DB 23; Length 884;  
Best Local Similarity 25.0%; Pred. No. 7.6;  
Matches 26; Conservative 23; Mismatches 46; Indels 9; Gaps 3;  
QY 34 SGFEGKDGAGYVNLKDTFKVPFKIE---EKKEENKPTFDVSK--KKDNPOVNHQ 88  
DB 14 SGQEVKNNTQTSLSKEEKLKREQLWLRQKQQQQQQEQEVQNKAKTSTNTS 73  
QY 89 LNESHRKEDLQREE---HSQKSDTKDVTATVLDKNNISSKST 128  
DB 74 TEEVKKSQRQRIEWEKRLQKQKQATKEKTTTITIKKVSQTT 117  
RESULT 25  
AAG08296  
ID AAG08296 standard; Protein; 329 AA.  
XX  
AC AAG08296;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 5775.  
DE Arabidopsis thaliana  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128713.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134258.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144634.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149315.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.

Query Match	12.2%; Score 85; DB 21; Length 329;	
Best Local Similarity	23.8%; Pred. No. 2.2;	
Matches 24; Conservative 22; Mismatches 33; Indels 22; Gaps 4;		
Qy	16 PVYKGELEKGYQDFGWEISGFEKKGADGYVINLSKDTFIKVPVKKIEKKEENKPTFDV 75	
Db	213 PIMERSLKRHVAEADWDQVD-----INLKVGGQQQDV--KLEKEKEEKQ--DM 257	
Qy	76 SKKDNPNVNSQLNESHKEDLQREHSQKSDSTKDVAT 116	
Db	258 SNEEDEEE-----EEEEKQDMSEDDKEEEDQEEREKT 291	
RESULT 26		
AA0808295		
ID	AA0808295 standard; Protein; 339 AA.	
XX		
AC	AA0808295;	
XX		
DT	17-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 5774.	
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX		
OS	Arabidopsis thaliana.	
XX		
FN	EPI033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0129845.	
PR	19-APR-1999; 99US-0130077.	
PR	21-APR-1999; 99US-0130449.	
PR	23-APR-1999; 99US-0130510.	
PR	28-APR-1999; 99US-0130891.	
PR	99US-0131449.	
PR	30-APR-1999; 99US-0132048.	
PR	30-APR-1999; 99US-0132407.	
PR	04-MAY-1999; 99US-0132484.	
PR	05-MAY-1999; 99US-0132485.	
PR	06-MAY-1999; 99US-0132486.	
PR	07-MAY-1999; 99US-0132487.	
PR	06-MAY-1999; 99US-0132487.	
PR	11-MAY-1999; 99US-0134256.	
PR	14-MAY-1999; 99US-0134219.	
PR	14-MAY-1999; 99US-0134219.	
PR	14-MAY-1999; 99US-0134221.	
PR	14-MAY-1999; 99US-0134370.	
PR	18-MAY-1999; 99US-0134768.	
PR	19-MAY-1999; 99US-0134941.	
PR	20-MAY-1999; 99US-0135124.	
PR	21-MAY-1999; 99US-0135353.	
PR	24-MAY-1999; 99US-0135629.	
PR	25-MAY-1999; 99US-0136021.	
PR	27-MAY-1999; 99US-0136332.	
PR	28-MAY-1999; 99US-0136782.	
PR	01-JUN-1999; 99US-0137222.	
PR	03-JUN-1999; 99US-0137528.	
PR	04-JUN-1999; 99US-0137502.	
PR	07-JUN-1999; 99US-0137724.	
PR	08-JUN-1999; 99US-0138094.	
PR	10-JUN-1999; 99US-0138540.	
PR	10-JUN-1999; 99US-0138847.	
PR	14-JUN-1999; 99US-0139119.	
PR	16-JUN-1999; 99US-0139452.	
PR	16-JUN-1999; 99US-0139453.	
PR	17-JUN-1999; 99US-0139492.	
PR	18-JUN-1999; 99US-0139454.	
PR	18-JUN-1999; 99US-0139455.	
PR	18-JUN-1999; 99US-0139456.	
PR	18-JUN-1999; 99US-0139457.	
PR	18-JUN-1999; 99US-0139458.	
PR	18-JUN-1999; 99US-0139459.	
PR	18-JUN-1999; 99US-0139460.	
PR	18-JUN-1999; 99US-0139461.	
PR	18-JUN-1999; 99US-0139462.	
PR	18-JUN-1999; 99US-0139463.	
PR	18-JUN-1999; 99US-0139750.	
PR	18-JUN-1999; 99US-0139763.	
PR	18-JUN-1999; 99US-0139817.	
PR	22-JUN-1999; 99US-0139899.	
PR	23-JUN-1999; 99US-0140353.	
PR	23-JUN-1999; 99US-0140354.	
PR	24-JUN-1999; 99US-0140693.	
PR	28-JUN-1999; 99US-0140823.	
PR	29-JUN-1999; 99US-0140	

PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145951.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	08-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149829.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145951.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	0	

PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.

```
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.2%; Score 85; DB 21; Length 358;
Best Local Similarity 23.8%; Pred. No. 2.5;
Matches 24; Conservative 22; Mismatches 33; Indels 22; Gaps 4;

QY 16 PVYGELEKGYQFDQWEISGPEGKDGAVYINLSKDTFKVPFKIEEKEEENKPTFDV 75
DB 242 PIMERSLKRHVEADWDQVD-----INLKVGGQDDV--KLEEKEEKEEQ--DM 286

QY 76 SKKDNQVNHSQLNSHRKEDLQREHSQKSDSTKDVAT 116
DB 287 SNEEDEEE-----EEEEKQDSEDDKEEDEQEEREKT 320

RESULT 28
ABP39011
ID ABP39011 standard; Protein; 465 AA.
AC ABP39011;
XX
XX
DT 24-JUL-2002 (first entry)
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3856.
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
PD 30-APR-2002.
XX
XX 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
XX
XX N-PSDB; ABN91556.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS Disclosure; SEQ ID 3856; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences

CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 465 AA;

Query Match 12.2%; Score 85; DB 23; Length 465;
Best Local Similarity 32.6%; Pred. No. 3.5;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY 61 IEEKKEEENKPTFDVSKKKNQVNHSQLNSHRKEDLQ-----REEHSQKSDS--- 109
DB 3 MEENKNQPNKE--NMSNKDDNA---THLNDSHRNEDLELFRNKNARRRRRIDNOSK 56

QY 110 TKDVTAT-----VLDKNNISSKSTTNPNK 134
DB 57 EXDATSTOSQLETPMDKFDINHKS--HNQNK 86

RESULT 29
AAG81782
ID AAG81782 standard; Protein; 472 AA.
XX
AC AAG81782;
XX
XX 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:658.
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX
XX Kimmerly WJ;
XX
XX WPI; 2001-316495/33.
XX
XX N-PSDB; AAH52632.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 18; Page 210; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
```

CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the preparation of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.

```

AA      SQ      Sequence      472 AA;

Query Match          12.2%; Score 85; DB 22; Length 472;
Best Local Similarity 32.6%; Pred. No. 3.6;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

Qy      61 IEKKKEENKPTFDVSKKKONPQVNSQLNESHKEDLQ-----REEHSQKSQS--- 109
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db      3 MEENKNQPKE--NWSNKDDNA----THLNDSHRNEDELEFRRKNKNARQRERRRIDNQSK 56

Qy      110 TKQVTAT-----VLDKNNISSKSTTNNPNK 134

Db      57 EKATSTOSOLETKPMDFLDNHKS--HNQNK 86

```

RESULT 30  
AAG82482  
ID AAG82482 standard; Protein; 720 AA.  
XX  
XX AC  
XX AAG82482;  
XX  
XX  
DT 03-SEP-2001 (first entry)

XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:2058.

Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination; endocarditis.

XX Staphylococcus epidermidis.

XX PN WO200134809-A2.

XX 17-MAY-2001 PD

09-NOV-2000; 2000WO-US30782.

AA 09-NOV-1999; 99US-0164258.

PA (GLAX) ) GLAXO GROUP LTD.

PI Kimmerly WJ;

XX  
DR WPI: 2001-316495/33.

DR N-PSDB; AAH53332.

Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
PT useful for vaccinating against infections, e.g. endocarditis -

PS Claim 18; Page 558; 2188pp; English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*. (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the *S. epidermidis* polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to AAH55090 represent specifically claimed *S. epidermidis* genomic DNA polynucleotide sequences from the present invention. AAH55091 to AAH55098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention.

N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464.

XX	Sequence	720 AA;
----	----------	---------

Query Match: 12.2%; Score 85; DB 22; Length 720;  
Best Local Similarity 28.0%; Pred. No. 6.4;  
Matches 23; Conservative 15; Mismatches 36; Indels

Qy	61	I E E K K E E N K P T F D V S K K C O -----N P Q W N I S Q L M E S H R K E D L O R E B H S Q K S D S T K-----D 112
		: : : : :
D b	38	L E E G I K A L D K K F K A S Q A K D T N K Q T Q N N H Q K N N K Q N S N D K E Q O S J O N N S K P T K K X E Q N 97
		: : : : :
Qy	113	V T A T V L D K N N I S S K S T T N N P N K 134
		: : : : :
D b	98	N K G K O O N K N K T N K N K O K N K N K 119
		: : : : :

RESULT 31  
ABP40123  
ID ABP40123 standard; Protein: 728 AA.

XX  
AC ABP40123;

24-JUL-2002 (first entry)

Staphylococcus epidermidis

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW

XX  
KW antibakteriell; keine Therapie.XX  
XXVXX  
XX  
11-01-2000 17:17

FD  
XX  
50-HFR-2002.

XX  
13-AUG-1958, 2000-VIS-001

PR 08-NOV-1997; 97US-064964P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;

DR WPI; 2002-381255/41.

XX  
N-ESSE, ADM22000

PS Disclosure; SEQ ID 4968; 267pp; English.

ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP3512 to ABP3960. The *S. epidermidis* sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly *S. epidermidis* infections. The sequences can be used to screen for compounds able to interfere with the *S. epidermidis* life cycle or inhibit *S. epidermidis* infection.

N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO web site.

AA		
SQ	Sequence	728 AA;

Query Match 12.2%; Score 85; DB 23; Length 728;  
Best Local Similarity 28.0%; Pred. No. 6.5;



Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 61 IEEKKEENKPTFDVSKKD-----NPQVNHSQLNESHKEDLQREHSQKSDSTK-----D 112

DB 46 LEEQIKALDKKFKASQAKDTNKQNTQNNHQKSNKQNSDKKQKQKSKPTKKEQN 105

QY 113 VTATVLDKNNISSKSTTNNPNK 134

DB 106 NKGQKQNNKNTKNNKQNNK 127

## RESULT 32

AAB18268  
ID AAB18268 standard; Protein; 1166 AA.

XX AC AAB18268;

XX DT 07-NOV-2000 (first entry)

XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:125.

XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
XX KW antimalarial; malaria; protozoacide; infection; insecticide.

XX OS Plasmodium falciparum.

XX PN WO200025728-A2.

XX PD 11-MAY-2000.

XX PF 05-NOV-1999; 99WO-US26796.

XX PR 05-NOV-1998; 98US-0107131.

XX PA (HOFF/) HOFFMAN S.

XX PA (CARU/) CARUCCI D.

XX PA (GARD/) GARDNER M.

XX PA (VENT/) VENTER J C.

XX PI Hoffman S, Carucci D, Gardner M, Venter JC;

XX DR WPI; 2000-365347/31.

XX PT Proteins encoded by chromosome 2 of the human malarial parasite,  
XX PT plasmodium falciparum, useful as antimalarial vaccines and in the  
XX PT diagnosis of P.falciparum infection -

XX PS Disclosure; Page 293-296; 577pp; English.

XX CC The present invention describes proteins and their fragments (I) encoded  
XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
XX CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
XX CC vaccines against P. falciparum infection comprising (I) or (II).  
XX CC (I) and (II) are useful for the development of vaccines against  
XX CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
XX CC antibody raised to immunogens comprising the sequences of (I), are  
XX CC useful in the detection of infection with P. falciparum. Furthermore,  
XX CC (I) (especially when they are refined or secreted or membrane proteins)  
XX CC can aid the identification of drugs to treat or prevent P. falciparum  
XX CC infection, or they can be used to identify drug resistance in  
XX CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
XX CC subsequent identification of proteins encoded by it will help to expand  
XX CC our understanding of parasite biology, a process hampered by the  
XX CC complexity of the parasitic lifecycle, and provide new targets for  
XX CC vaccine and drug development. Parasite resistance to drugs and mosquito  
XX CC resistance to insecticides have led to a resurgence of malaria in many  
XX CC parts of the world, and there is a pressing need for vaccines and new  
XX CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
XX CC and protein sequences given in the present invention, but which are not  
XX CC specifically mentioned within the specification.

XX SQ Sequence 1166 AA;

## Query Match

Best Local Similarity 12.2%; Score 85; DB 21; Length 1166;

Matches 28; Conservative 16; Mismatches 34; Indels 20; Gaps 3;

QY 38 GKQDAGVYINLSKDTFTFKPVFKKIEEKEENKPTFDVSKKKNPQVNHSQLNESHKED 97

DB 614 GKKEE-----KCVTKIIEETKNEEMEPNQEQDKDN-----ELKEKNDKEE 655

QY 98 L--QREHSQKSDSTKDVATVLDKNNISSKSTTNNPN 133

DB 656 KNDQEKNDQEKNDKEKNDKEKNDQEKNDQEKV 693

## RESULT 33

AAW69361  
ID AAW69361 standard; Protein; 1978 AA.

XX AC AAW69361;

XX DT 01-DEC-1998 (first entry)

XX DE Tetradotoxin-sensitive sodium channel PN4.

XX KW Tetradotoxin-sensitive sodium channel; rat; PN4 sodium channel; stroke;

XX KW nervous system disorder; epilepsy; brain injury; diabetic neuropathy;

XX KW AIDS-associated neuropathy; therapy.

XX OS Rattus sp.

XX PN WO9838302-A2.

XX PD 03-SEP-1998.

XX PF 20-FEB-1998; 98WO-EP00997.

XX PR 26-FEB-1997; 97US-0039447.

XX PA (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX PI Delgado SG, Dietrich PS, Fish LM, Herman RC, Sangameswaran L;

XX DR WPI; 1998-481204/41.

XX DR N-ESDB; AAV58420.

XX PT New rat tetradotoxin-sensitive sodium channel alpha subunit and DNA

XX PT - for detecting inhibitors which alleviate pain, and treating

XX PT nervous system disorders, e.g. epilepsy, stroke, diabetic and AIDS

XX PT neuropathy

XX PS Claim 16; Page 36-43; 87pp; English.

XX CC This sequence is encoded by the rat PN4 sodium channel cDNA clone of the  
XX CC invention. The DNA sequence was isolated from a peripheral nerve from  
XX CC a rat dorsal ganglia. The PN4 sodium channel sequences are  
XX CC tetradotoxin-sensitive sodium channels. The protein is used in assays for  
XX CC detecting inhibitors of tetradotoxin-sensitive sodium channels, which  
XX CC alleviate pain. The probes can be used to detect and isolate the DNA or  
XX CC protein in tissues. The antibodies can also be used to isolate the  
XX CC protein. The protein is used as a therapeutic target for compounds to  
XX CC treat disorders of the nervous system, such as epilepsy, stroke and brain  
XX CC injury, diabetic neuropathy, and AIDS-associated neuropathy, etc.

XX SQ Sequence 1978 AA;

Query Match 12.2%; Score 85; DB 19; Length 1978;

Best Local Similarity 25.2%; Pred. No. 26;

Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

QY 27 QFDCWEISGPEGKDGAGVINLSKDTFTFKPVFKKIEEKEENKPTFDVSKKKNPQVNH 86

DB 991 EMNNIQISIRIKKGVAVT-KVKVHAFMQAHFK---QREADEVPDLDELYEKANCIAH 1046

QY 87 SQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPN 133





PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143257.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145219.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 10-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 12.1%; Score 84.5; DB 21; Length 309;

Best Local Similarity 24.8%; Pred. No. 2.3;

Matches 32; Conservative 16; Mismatches 46; Indels 35; Gaps 5;

QY 15 LPVYKGELEKGYQFDGWELSGPEGKK-----DAGYVINLS-----KDTFIKPVFKKI 61  
Db 141 LPVYK-----DNDSFTVKRISDLQASKESPARAVASDKGTFVPSSSQEKDPVAKTILNKV 196  
QY 62 EEKKEENKPTFDVSKKKNPQVN-----HSQLNESHKKE-----DLOREBH 103  
Db 197 KEKSDDEGKPLPAEFDGKENPVANSKNDKSDSGVRSRERERGRGRDARGRDSRDR 256  
QY 104 SOKSDSTKD 112  
Db 257 GRESDRERD 265

RESULT 38  
AAG35427  
ID AAG35427 standard; Protein; 361 AA.

XX AAG35427; 18-JUN-1999; 99US-0139462.  
XX AC 18-JUN-1999; 99US-0139463.  
XX DT 18-JUN-1999; 99US-0139750.  
XX XX 18-JUN-1999; 99US-0139763.  
XX DE 21-JUN-1999; 99US-0139817.  
XX XX 22-JUN-1999; 99US-0139899.  
XX KW 23-JUN-1999; 99US-0140353.  
XX KW 24-JUN-1999; 99US-0140354.  
XX KW 24-JUN-1999; 99US-0140695.  
XX XX 28-JUN-1999; 99US-0140823.  
XX OS 29-JUN-1999; 99US-0140991.  
XX XX 30-JUN-1999; 99US-0141287.  
XX PN 01-JUL-1999; 99US-0141842.  
XX XX 01-JUL-1999; 99US-0142154.  
XX XX 02-JUL-1999; 99US-0142055.  
XX PD 06-JUL-1999; 99US-0142390.  
XX XX 08-JUL-1999; 99US-0142803.  
XX PF 09-JUL-1999; 99US-0142920.  
XX XX 12-JUL-1999; 99US-0142977.  
XX XX 13-JUL-1999; 99US-0143542.  
XX XX 14-JUL-1999; 99US-0143624.  
XX XX 15-JUL-1999; 99US-0144005.  
XX XX 16-JUL-1999; 99US-0144085.  
XX XX 16-JUL-1999; 99US-0144086.  
XX XX 19-JUL-1999; 99US-0144325.  
XX XX 19-JUL-1999; 99US-0144331.  
XX XX 19-JUL-1999; 99US-0144332.  
XX XX 19-JUL-1999; 99US-0144333.  
XX XX 19-JUL-1999; 99US-0144334.  
XX XX 19-JUL-1999; 99US-0144335.  
XX XX 20-JUL-1999; 99US-0144352.  
XX XX 20-JUL-1999; 99US-0144632.  
XX XX 20-JUL-1999; 99US-0144884.  
XX XX 21-JUL-1999; 99US-0144814.  
XX XX 21-JUL-1999; 99US-0145086.  
XX XX 21-JUL-1999; 99US-0145088.  
XX XX 22-JUL-1999; 99US-0145085.  
XX XX 22-JUL-1999; 99US-0145087.  
XX XX 22-JUL-1999; 99US-0145089.  
XX XX 22-JUL-1999; 99US-0145192.  
XX XX 23-JUL-1999; 99US-0145145.  
XX XX 23-JUL-1999; 99US-0145218.  
XX XX 23-JUL-1999; 99US-0145224.  
XX XX 26-JUL-1999; 99US-0145276.  
XX XX 27-JUL-1999; 99US-0145913.  
XX XX 27-JUL-1999; 99US-0145918.  
XX XX 27-JUL-1999; 99US-0145919.  
XX XX 28-JUL-1999; 99US-0145951.  
XX XX 02-AUG-1999; 99US-0146386.  
XX XX 02-AUG-1999; 99US-0146388.  
XX XX 02-AUG-1999; 99US-0146389.  
XX XX 03-AUG-1999; 99US-0147038.  
XX XX 04-AUG-1999; 99US-0147204.  
XX XX 04-AUG-1999; 99US-0147302.  
XX XX 05-AUG-1999; 99US-0147192.  
XX XX 05-AUG-1999; 99US-0147260.  
XX XX 06-AUG-1999; 99US-0147303.  
XX XX 06-AUG-1999; 99US-0147416.  
XX XX 09-AUG-1999; 99US-0147493.  
XX XX 09-AUG-1999; 99US-0147935.  
XX XX 10-AUG-1999; 99US-0148171.  
XX XX 11-AUG-1999; 99US-0148319.  
XX XX 12-AUG-1999; 99US-0148341.  
XX XX 13-AUG-1999; 99US-0148565.  
XX XX 13-AUG-1999; 99US-0148684.  
XX XX 16-AUG-1999; 99US-0149368.  
XX XX 17-AUG-1999; 99US-0149175.  
XX XX 18-AUG-1999; 99US-0149426.  
XX XX 20-AUG-1999; 99US-0149722.  
XX XX 20-AUG-1999; 99US-0149723.  
XX XX 20-AUG-1999; 99US-0149929.  
XX XX 23-AUG-1999; 99US-0149902.  
XX XX 23-AUG-1999; 99US-0149930.  
  
XX AAG35427; 25-FEB-2000; 2000EP-0301439.  
XX AC 25-FEB-1999; 99US-0121825.  
XX DT 05-MAR-1999; 99US-0123180.  
XX XX 09-MAR-1999; 99US-0123548.  
XX XX 23-MAR-1999; 99US-0125788.  
XX XX 25-MAR-1999; 99US-0126264.  
XX XX 29-MAR-1999; 99US-0126785.  
XX KW 01-APR-1999; 99US-0127462.  
XX KW 06-APR-1999; 99US-0128234.  
XX XX 08-APR-1999; 99US-0128714.  
XX OS 16-APR-1999; 99US-0129845.  
XX XX 19-APR-1999; 99US-0130077.  
XX PN 21-APR-1999; 99US-0130449.  
XX XX 23-APR-1999; 99US-0130510.  
XX XX 23-APR-1999; 99US-0130891.  
XX XX 28-APR-1999; 99US-0131449.  
XX XX 30-APR-1999; 99US-0132048.  
XX XX 30-APR-1999; 99US-0132407.  
XX XX 04-MAY-1999; 99US-0132484.  
XX XX 05-MAY-1999; 99US-0132485.  
XX XX 06-MAY-1999; 99US-0132486.  
XX XX 06-MAY-1999; 99US-0132487.  
XX XX 07-MAY-1999; 99US-0132863.  
XX XX 11-MAY-1999; 99US-0134256.  
XX XX 14-MAY-1999; 99US-0134218.  
XX XX 14-MAY-1999; 99US-0134219.  
XX XX 14-MAY-1999; 99US-0134221.  
XX XX 14-MAY-1999; 99US-0134370.  
XX XX 18-MAY-1999; 99US-0134768.  
XX XX 19-MAY-1999; 99US-0134941.  
XX XX 20-MAY-1999; 99US-0135124.  
XX XX 21-MAY-1999; 99US-0135353.  
XX XX 24-MAY-1999; 99US-0135629.  
XX XX 25-MAY-1999; 99US-0136021.  
XX XX 27-MAY-1999; 99US-0136392.  
XX XX 28-MAY-1999; 99US-0136782.  
XX XX 01-JUN-1999; 99US-0137222.  
XX XX 03-JUN-1999; 99US-0137528.  
XX XX 04-JUN-1999; 99US-0137502.  
XX XX 07-JUN-1999; 99US-0137724.  
XX XX 08-JUN-1999; 99US-0138094.  
XX XX 10-JUN-1999; 99US-0138540.  
XX XX 10-JUN-1999; 99US-0138847.  
XX XX 14-JUN-1999; 99US-0139119.  
XX XX 16-JUN-1999; 99US-0139452.  
XX XX 16-JUN-1999; 99US-0139453.  
XX XX 17-JUN-1999; 99US-0139492.  
XX XX 18-JUN-1999; 99US-0139454.  
XX XX 18-JUN-1999; 99US-0139455.  
XX XX 18-JUN-1999; 99US-0139456.  
XX XX 18-JUN-1999; 99US-0139457.  
XX XX 18-JUN-1999; 99US-0139458.  
XX XX 18-JUN-1999; 99US-0139459.  
XX XX 18-JUN-1999; 99US-0139460.  
XX XX 18-JUN-1999; 99US-0139461.

PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 08-OCT-1999; 99US-0158029.  
PR 12-OCT-1999; 99US-0158232.  
PR 13-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 12.1%; Score 84.5; DB 21; Length 361;  
Best Local Similarity 24.8%; Pred. No. 2.8;  
Matches 32; Conservative 16; Mismatches 46; Indels 35; Gaps 5;

QY 15 LPVYKGELEKGYQPDGWEISGFEGKK-----DAGVYNLS-----KDTFKPVFKKI 61  
Db 193 IPVYK-----DNDSTFTVKRISDLQAKESPARAVASDKGTVPVSSSSQEKDPVAKTILNKV 248  
QY 62 EKKKEENKPTFDVSKKDNFQVN-----HSQLNESHKKE-----DLOREH 103  
Db 249 KEKSDDEGKPLPAEFDGKENPVANSKNDKSDSGVDRSRERSRGRERDARGDRSDRDSR 308

QY 104 SKSDSTKD 112

Db 309 GRESDRERD 317

RESULT 39  
AAB35803

ID AAB35803 standard; Protein; 390 AA.  
XX  
AC AAB35803;  
XX  
DT 23-FEB-2001 (first entry)  
XX  
DE Protein involved in cell cycle regulation SEQ ID 38.  
XX  
KW Cell cycle regulation; corn; transgenic plant; cyclin; maize; soybean;  
KW cyclin-dependent kinase; sunflower; sorghum; canola; wheat; alfalfa;  
XX cotton; rice; barley; millet.  
OS Zea mays.  
XX  
FN WO200065040-A2.  
XX  
PD 02-NOV-2000.  
XX  
PF 13-APR-2000; 2000WO-US09975.  
XX  
PR 22-APR-1999; 99US-0130849.  
XX (PION-) PIONEER HI-BRED INT INC.  
XX Helentjaris TG, Habben JE, Sun Y;  
XX  
DR WPI; 2000-687333/67.  
DR N-PSDB; AAC83110.  
XX  
PT Nucleic acids useful for producing transgenic plants, preferably maize,  
PT with increased cell cycle gene activity, preferably activity of cyclin  
PT and/or cyclin-dependent kinase -  
XX  
PS Claim 16; Page 111-112; 122pp; English.  
XX  
CC Polynucleotide sequences AAC83101 - AAC83113 encode proteins AAB35794 -  
CC AAB35806 which are involved in regulating the cell cycle. The protein and  
CC DNA sequences have been isolated from Zea mays (corn), and the invention  
CC also includes oligonucleotides AAC83114 - AAC83139 which are related to  
CC the cell cycle polynucleotides. The cell cycle polynucleotide sequences  
CC are useful for producing transgenic plants such as maize, soybean,  
CC sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley and  
CC millet with increased levels of cell cycle gene activity, such as  
CC activity of cyclin and cyclin-dependent kinases. The DNA sequences are  
CC also useful as probes for detecting deficiencies in the level of mRNA in  
CC screening for desired transgenic plants, for detecting mutations in the  
CC gene, for monitoring upregulation of expression or changes in enzyme  
CC activity in screening assays of compounds, for detecting any number of  
CC allelic variants, orthologs or paralogues of the gene, and site-directed  
CC mutagenesis in eukaryotic cells. The DNA sequences are also useful for  
CC recombinant expression of the encoded polypeptides and as immunogens for  
CC preparing and screening antibodies. A transgenic plant comprising an  
CC expression cassette including a cell cycle regulatory gene is useful for  
CC assaying enzyme agonists and antagonists, and as immunogens or antigens  
CC to obtain antibodies. The antibodies are useful in assaying expression  
CC levels of cell cycle regulatory proteins, for identifying and isolating  
CC nucleic acids from expression libraries, for identifying homologues of  
CC polypeptides from other species, and for purification of the proteins.  
XX  
SQ Sequence 390 AA;

Query Match 12.1%; Score 84.5; DB 21; Length 390;  
Best Local Similarity 24.8%; Pred. No. 3.1;  
Matches 32; Conservative 16; Mismatches 46; Indels 35; Gaps 5;

QY 15 LPVYKGELEKGYQPDGWEISGFEGKK-----DAGVYNLS-----KDTFKPVFKKI 61  
Db 248 IPVYK-----DNDSTFTVKRISDLQAKESPARAVASDKGTVPVSSSSQEKDPVAKTILNKV 303  
QY 62 EKKKEENKPTFDVSKKDNFQVN-----HSQLNESHKKE-----DLOREH 103  
Db 304 KEKSDDEGKPLPAEFDGKENPVANSKNDKSDSGVDRSRERSRGRERDARGDRSDRDSR 363

QY 104 SOKSDSTKD 112  
Db 364 GRESDRERD 372  
RESULT 40  
AAG35426  
ID AAG35426 standard; Protein; 416 AA.  
XX  
AC AAG35426;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 43273.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX  
OS Zea mays subsp. mays.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 12.1%; Score 84.5; DB 21; Length 416;  
Best Local Similarity 24.8%; Pred. No. 3.4;  
Matches 32; Conservative 16; Mismatches 46; Indels 35; Gaps 5;  
QY 15 LPVYKELEKGVQFDWEISGPEGK-----DAGYVINLS-----KDTFKPVFKKI 61  
DB 248 IPVYK----DNDSFTVKRISDLQAKSPARAVASDKGTFVPSSSQEKDPVAKTILNKV 303  
QY 62 EEKKEENKPTFDVSKKKNDPNQW-----HSOLNESHKKE-----DLQREEH 103  
DB 304 KEKSDGEGPLPAEFGKENVANSKNDKSDGVDNRERERSRGRERDARGDSDRDR 363

QY 104 SQKSDSTKD 112  
DB 364 GRESDRERD 372

Search completed: February 10, 2004, 10:53:52  
Job time : 31.6275 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 9.67557 Seconds  
(without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSSSTIVSEEDFILPVYK.....ATVLDKNISSKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	696	100.0	2140	2 F95074	serine proteinase,
2	693	99.6	2144	2 A37942	metalloproteinase,
3	110	15.8	558	2 T18467	hypothetical prote
4	108.5	15.6	1038	2 J05497	claustrin - chicke
5	99	14.2	211	2 T25911	hypothetical prote
6	98	14.1	665	2 B71609	hypothetical prote
7	97.5	14.0	1345	2 S46817	hypothetical prote
8	95	13.6	348	2 I37271	cyclicin II - human
9	95	13.6	622	2 A90570	lipoprotein (impor
10	94.5	13.6	312	2 G81339	probable membrane
11	94	13.5	210	2 T28771	hypothetical prote
12	94	13.5	535	2 T37189	hypothetical prote
13	93	13.4	1397	2 T10466	DNA topoisomerase
14	92.5	13.3	219	2 B72291	hypothetical prote
15	92.5	13.3	325	2 T18283	hypothetical prote
16	91	13.1	253	2 T32879	hypothetical prote
17	89.5	12.9	508	2 B81594	hypothetical prote
18	89.5	12.9	508	2 C72074	hypothetical prote
19	89.5	12.9	508	2 E86549	hypothetical prote
20	89	12.8	528	2 E36795	unknown protein F2
21	89	12.8	1888	2 T39009	hypothetical prote
22	89	12.8	3724	2 T18427	hypothetical prote
23	88.5	12.7	301	2 T33068	hypothetical prote
24	88.5	12.7	371	2 A71683	hypothetical prote
25	88.5	12.7	385	2 T20410	hypothetical prote
26	88.5	12.7	644	2 D86432	hypothetical prote
27	88.5	12.7	644	2 T47835	hypothetical prote
28	88.5	12.7	762	2 G88436	protein T04A8.13 [
29	88.5	12.7	791	2 T24435	hypothetical prote

ALIGNMENTS

RESULT 1

F95074

serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: F95074

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle,

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: F95074

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2140 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK74791.1; PID:g14972117; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0641

Query Match 100.0%; Score 696; DB 2; Length 2140;

Best Local Similarity 100.0%; Pred. No. 2.2e-46;

Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVNLKSDTFIKVPFKK 60

Db 1973 KEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVNLKSDTFIKVPFKK 2032

Qy 61 IEKKKEENKPTFDVSKKONPQVNHSQLNESHKRDLOREHSQKSDSTKDVATVLDK 120

Db 2033 IEKKKEENKPTFDVSKKONPQVNHSQLNESHKRDLOREHSQKSDSTKDVATVLDK 2092

Qy 121 NNISKSSTNNPNK 134

Db 2093 NNISKSSTNNPNK 2106

RESULT 2

A97942

metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C:Accession: A97942

R:Hoekings, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A97942  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99365.1; PID:g15458138; GSPDB:GN00174  
C:Genetics:  
A:Gene: pRTA  
C:Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 693; DB 2; Length 2144;  
Best Local Similarity 99.3%; Pred. No. 3.8e-46;  
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFK 60  
DB 1977 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFK 2036

QY 61 IEKKEEENKPTFDVSKKNDPQVNHSQLNESHREDLQREHSQKSDSTKDVATVLDK 120  
DB 2037 IEKKEEENKPTFDVSKKNDPQVNHSQLNESHREDLQREHSQKSDSTKDVATVLDK 2096

QY 121 NNISKSTTNPNK 134  
DB 2097 NNISKSTTNPNK 2110

RESULT 3  
T18467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2002  
C:Accession: T18467  
R:Lawson, D.; Bowman, S.; Barrell, B.  
A:Reference number: Z18937  
A:Accession: T18467  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-558 <LAW>  
A:Cross-references: EMBL:AL008970; NID:e1407852; PIDN:CAA15610.1  
C:Genetics:  
A:Map position: 3  
A:Introns: 84/1; 160/1  
A:Note: C0465c

Query Match 15.8%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 0.31;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 7 IVSEEDFILPVY-----KGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPV 57  
DB 60 ILGFEDDILYEYCISQLKQSKK--KADGEEDKYLNAKKLKLNLGFIGNKKSDIFBEL 117

QY 58 FKKI--EKKKEE-----ENKPTFDVSK-KKNDPQVNHSQLNE-----SHRK 95  
DB 118 DELLINEEKKEEHIADTLNENK-TNDIRKVRNENINENYVNEKDISNKKDKEHVSHQN 176

QY 96 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISSKSTTN 130  
DB 177 EHNINNVNLKKEKEYTDIQDKRKHRSLSQKSDSYKKRPFNKRKTSIER-SLSNRYDE 235

QY 131 NPNK 134  
DB 236 KTNK 239

RESULT 4  
JC5497  
claustrin - chicken  
N:Alternate names: keratan sulfate proteoglycan  
C:Species: Gallus gallus (chicken)  
C:Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 21-Jul-2000  
C:Accession: JC5497; PC4334; S37561

R:Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A:Title: Claustrian, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JC5497; MUID:94157526; PMID:7906711  
A:Accession: JC5497  
A:Molecule type: mRNA  
A:Residues: 1-1038 <BUR1>  
A:Cross-references: EMBL:X67778; NID:g406318; PIDN:CAA47988.1; PID:g406319  
A:Accession: PC4334  
A:Molecule type: protein  
A:Residues: 79-83;299-412;485-502 <BUR2>  
A:Experimental source: brain  
C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nerve  
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F:267-270/Region: cell attachment (R-G-D) motif  
F:112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 15.6%; Score 108.5; DB 2; Length 1038;  
Best Local Similarity 28.6%; Pred. No. 0.83;  
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFK- 59  
DB 595 KPETKTIVAEDV-----TTKEEQLGKSETSEKQASEKQDVKPKVTKESVKKEVKA 646

QY 60 KIEKKEEENKPTFDVSKKNDPQVNHSQLNESHREDLQRE-----EHSQKSDSTKDV 113  
DB 647 KPEEKKDEKPKPKESVKKEKPLI---KKEPKKEDIKKVEKVKKEKKEKKEKKEV 702

RESULT 5  
T25911  
hypothetical protein T23B3.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25911  
R:Maggi, L.; Le, T.  
submitted to the EMBL Data Library, February 1997  
A:Description: The sequence of C. elegans cosmid T23B3.  
A:Reference number: Z20109  
A:Accession: T25911  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-211 <MAG>  
A:Cross-references: EMBL:U88309; PIDN:AAB42334.1; GSPDB:GN00019; CESP:T23B3.5  
A:Experimental source: strain Bristol N2; clone T23B3  
C:Genetics:  
A:Gene: CESP:T23B3.5  
A:Map position: 1  
A:Introns: 30/2; 200/3

Query Match 14.2%; Score 99; DB 2; Length 211;  
Best Local Similarity 29.4%; Pred. No. 0.76;  
Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

QY 37 EGKGDAGYVINLSKDTFIKPVFKIEKKEEENKPTFDVSKKNDPQVNH-----SQLNES 92  
DB 69 EGKGDGKKEKKGDKGDKKEEKKDEKKDKGDKKEDDKKEDDKKEDDKKEDDKKEDDK 128

QY 93 HRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 134  
DB 129 EKKDDKDKEDKDKEDKDKEDKDKESKSKSKSKSKSKSKSKSKSK 170

RESULT 6  
B71609  
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: B71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Partea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A;Reference number: A71600; MUID:59021743; PMID:9804551

A;Accession: B71609

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-665 <GAR>

A;Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AACT71925.1; PID:g384524

A;Experimental source: clone 3D7

C;Genetics:

A;Gene: PFB0680W

```
Query Match      14.1%; Score 98; DB 2; Length 665;
Best Local Similarity 27.1%; Pred. No. 3.3;
Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;

QY 23 EKGYPFGWEI--SGPFGKKDAGVINLSKDTFTKPVFKIEEKK-----EENK 70
Db 158 ERGKQ-----DISNSAENKKD-----VKEGVKELEKKKEEKSDDHKVEENK 201

QY 71 PTFD---VSKKDNPOVNSQLNESHKEDLQR-EHSHQSKSDSTKDVTTATVLDKNNISS 125
Db 202 KSDDHKVEENKSDDHKVEENKSDDHKIEEVKKVEEHEDEE-----DKKEKS 252

QY 126 KSTNNPNK 134
Db 253 ENKNKDENK 261
```

#### RESULT 7

S46817

C;Species: Saccharomyces cerevisiae

C;Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 19-Apr-2002

C;Accession: S46817

R;Favell, T.

submitted to the EMBL Data Library, June 1994

A;Description: The sequence of *S. cerevisiae* cosmid 9205.

A;Reference number: S46795

A;Accession: S46817

A;Molecule type: DNA

A;Residues: 1-1345 <FAV>

A;Cross-references: EMBL:U10556; NID:g500825; PIDN:AAB68895.1; PID:g500838; MIPS:YHR080C

C;Genetics:

A;Cross-references: SGD:S0001122

A;Map position: 8R

C;Keywords: transmembrane protein

```
Query Match      14.0%; Score 97.5; DB 2; Length 1345;
Best Local Similarity 27.0%; Pred. No. 8;
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY 19 KGELEKGYQFDGWEISGFEKK-DAGVINLSKDTFTKPVFKIEEKKKEENKFTPDVSK 77
Db 1109 KGATEKG-----SVEGQKVSVDYMLSELRII-----SRAKSKPKPVKKVMK 1149

QY 78 KDNPOVNSQLNESHKEDLQREHSHQSKSDSTKDVTTATVLD--KNNSKSTTN---NP 132
Db 1150 SHDKRPFPHSKVE-----QKSSGRKSDDNKDILTHILDFVQNNFSPSEIFNKLISP 1201

QY 133 NK 134
Db 1202 QK 1203
```

#### RESULT 8

I37271

C;Species: Homo sapiens (man)

C;Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000

C;Accession: I37271; S52774

R;Hees, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.

Exp. Cell Res. 218, 174-182, 1995

A;Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the

A;Reference number: I37271; MUID:95255491; PMID:7737358

A;Accession: I37271

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-348 <HES>

A;Cross-references: EMBL:Z46788; NID:g758586; PIDN:CAA86752.1; PID:g758587

Query Match 13.6%; Score 95; DB 2; Length 348;

Best Local Similarity 30.9%; Pred. No. 2.7;

Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

```
QY 19 KGELEKGYQFDGWEISGFEKKDAGVINLSKDTFTKPVFKIEEKKKEENKPTTF---DV 75
Db 205 ESEGEKG---GTEKDSKKGKDS-----KKGKDSAIELQAVKADKKDEKDKDANKGE 256

QY 76 SK--KDNPOVNSQLN-----ESHRKEDLQREHSHQSKSDTKD---VTATVLDKNNI 123
Db 257 SKDAKKDAKEIKGKKKKKKPSPSTDSKDDVKKR---SKDATKDAKKVAKKDKTEKESA 313

QY 124 SSK 126
Db 314 DSK 316
```

#### RESULT 9

A90570

C;Species: Mycoplasma pulmonis (strain UAB CTIP)

C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001

C;Accession: A90570

R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galissou, F.; Moszer, I.;

Nucleic Acids Res. 29, 2145-2153, 2001

A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis

A;Reference number: A99512; MUID:21267165; PMID:11353084

A;Accession: A90570

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-622 <KUR>

A;Cross-references: GB:AL445566; PID:g14089879; PIDN:CAC13638.1; GSPDB:GN00153

A;Experimental source: strain UAB CTIP

C;Genetics:

A;Gene: MYPU 4650

A;Genetic code: SGC3

```
Query Match      13.6%; Score 95; DB 2; Length 622;
Best Local Similarity 25.4%; Pred. No. 5.3;
Matches 34; Conservative 26; Mismatches 50; Indels 24; Gaps 5;
```

```
QY 8 VSEDEFILPVYKGELEKGYQFDGWEISGFEKKDAGVINLSKDTFTKPVFKIEEKEE 67
Db 85 LSEDNIIIFSLNIENNAKFOLDEF-----VSKDEKFKIKFQEIFSQTE 128

QY 68 ENKPTFDVSKKD-----NPOVNSQLNES--HRKEDLQREHSHQSKSDTKDVTTATVLDKN 121
Db 129 Q-KITDNISSEKDEKNPKDNENNNSSDQKNDLQKNNSDKLNDVQDEKANKENS 187

QY 122 -NLSKSTTNPNK 134
Db 188 SNDSKKNKNTNK 201
```

#### RESULT 10

G81339

C;Species: Campylobacter jejuni (strain NCTC 11168)

C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C;Accession: G81339

R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel

Nature 403, 665-668, 2000

A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A;Reference number: A81250; MUID:20150912; PMID:10688204

```

A;Residues: 1-535 <LEI>
A;Cross-references: EMBL:U49945; PIDN:AAC47924.1; GSPDB:GN00029; CESP:C02H7.1
A;Experimental source: strain Bristol N2; clone C02H7
C;Genetics:
A;Gene: CESP:C02H7.1
A;Map position: X
A;Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match      13.5%; Score 94; DB 2; Length 535;
Best Local Similarity 23.4%; Pred. No. 5.3;
Matches 33; Conservative 26; Mismatches 56; Indels 26; Gaps 4;

Qy      4  SSTIVSEEPILPVYKGELEKGYQFGWEISGFEKKDAGVYNLSKDTFIKPVFKKIEE 63
Db      86  AAKISGKD-----AEETNMQLMGTNATNSRNGTG-----EEKKKKVKVKEDK 132

Qy      64  KKEEENKPTPDVSKKNDNPVNHSSQLNESHKREDLQREHSSQKSDSTK-----DV 113
Db      133  KGDEEKST---TKRSSKKEETHEEKEKSEKKSAEEKEKKKKSSSSSKRHKSSDRSSEK 189

Qy      114  TATVLDKNINSSKSTTNPNK 134
Db      190  SSEKSSKEKKKEKSTTDEKPK 210

```

RESULT 13  
Tl0466  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - malaria parasite (Plasmodium falciparum)  
C/Species: Plasmodium falciparum  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000  
C/Accession: Tl0466  
R:/Cheesman, S.J.  
submitted to the EMBL Data Library, September 1995  
A/Reference number: Z17031  
A/Accession: Tl0466  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-1397 <CHE>  
A/Cross-references: EMBL:X79345; NID:g994807  
C/Genetics:  
A/Gene: TopoII  
A/Map position: 14  
C/Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hydrolyzing)  
C/Keywords: ATP; DNA binding; isomerase; nucleus  
Query Match 13.4% Score 93; DB 2; Length 1397;

	Matches	35;	Conservative	33;	Mismatches	48;	Indels	26;	Gaps	6;
Qy	8	VSEEDRIL--PVYKGLEKGYQFDGHEISGFEGKKDAGYVNLNLSKDTFFKPVFKKIE--	63							
Db	1141	VKQDYDLLSMPIFSLTLEK---VEDLLTQLKEKERLEIURNITVETMWLKDIEKVEAI	1197							
Qy	64	-----KKBEENKPTFDVSKKKNQPNVHSQLNESHKKEDLQREHSQ-----KSDSTK	111							
Db	1198	EFORNVLSNREESNK--FKVARKQ-----GPSMKMKKKKKKLLSDSEEGGDTSDSSE	1250							
Qy	112	DVTATVLDKNNISKSTNNPN	133							

RESULT 14  
B72291  
hypothetical protein - *Thermotoga maritima* (strain MSB8)  
C/Species: *Thermotoga maritima*  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C/Accession: B72291  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.  
Nature 399, 323-329, 1999  
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome se

Nature 399, 323-329, 1999

A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: B72291  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-219 <ARN>  
A;Cross-references: GB:AE001771; GB:AE000512; NID:94981678; PIDN:AAD36218.1; PID:9498168  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM1142

Query Match 13.3%; Score 92.5; DB 2; Length 219;  
Best Local Similarity 28.1%; Pred. No. 2.5;  
Matches 36; Conservative 24; Mismatches 35; Indels 33; Gaps 8;  
QY 20 GELEKGYQ--FDGWEISG-----FEGKDGAGYVIN-LSKDTFTKPVFKKIEEKEENK 70  
DB 89 GEESVSRDLFAGVGVRGTPTFFFKGEGGLGYDGVVDKDNFIK-ILKYVAQLKED-- 145  
QY 71 PTFDVSKKKNPQVNHSQLNESHKED----LQREHSQKSDS-----TKD 112  
DB 146 --FTYLLKDDPFVGEPLIIIF-KEDADFVLEKDNANVKDVTVPNEVRDRIVYVTDSPD 202  
QY 113 VTATVLDK 120  
DB 203 VAKTLOEK 210

## RESULT 15

T18283  
hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
C;Species: Dictyostelium discoideum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T18283  
R;Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh  
Genetics 148, 1117-1125, 1998  
A;Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1  
A;Reference number: 214684; MUID:98198836; PMID:9539429  
A;Accession: T18283  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-325 <RIE>  
A;Cross-references: EMBL:U00796; NID:92702254; PID:g2702258; PIDN:AAC18634.1  
C;Genetics:  
A;Introns: 85/1

Query Match 13.3%; Score 92.5; DB 2; Length 325;  
Best Local Similarity 23.2%; Pred. No. 4;  
Matches 38; Conservative 23; Mismatches 42; Indels 61; Gaps 8;  
QY 8 VSEEDFILPVYK-GEL--EKGYQFDGWEISGFEK-----KDAGYV-----I 46  
DB 72 IGDEKULFRISKNGELIVLNELEFDNFHFK--EGKHLKSNFNIKDSGYATNEIEI 129  
QY 47 NLSKDTFTKPV-----FKK-----IIEKKEENKPTFDV 75  
DB 130 FLESTCLCKEITATQTKNSYKKNINKLPEEEEEEEEEEEEEEEQEEVEKPTISE 189  
QY 76 SKKNDPQVNHSQLNESHK-----DLQREHSQKSDSTKD 112  
DB 190 EEEETPAVSEEEKEEEEEETPAVSEEEKEEQEEDKEK 233

## RESULT 16

T32879  
hypothetical protein C17F3.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T32879  
R;Gattung, S.; Scheet, P.  
submitted to the EMBL Data Library, January 1998  
A;Description: The sequence of C. elegans cosmid C17F3.  
A;Reference number: Z21420  
A;Accession: T32879

A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-253 <GAT>  
A;Cross-references: EMBL:AF043692; PIDN:AAB97531.1; GSPDB:GN00019; CESP:C17F3.3  
A;Experimental source: strain Bristol N2; clone C17F3  
C;Genetics:  
A;Gene: CESP:C17F3.3  
A;Map position: 1  
A;Introns: 41/1

Query Match 13.1%; Score 91; DB 2; Length 253;  
Best Local Similarity 32.6%; Pred. No. 3.9;  
Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;  
QY 56 PVFKKTEEKKE---EENKPTFDVSKK-----KDNPQVNHSQLNESHK---- 94  
DB 61 PVAPKVEEKEEKEEKKADDEKKEEKKKTEEDKKISVKVKTQETKSERKDKK 120  
QY 95 ---KEDLQREHSQKSDSTKDVTTATVLDKKNISK 126  
DB 121 DEKDEKKEEKEKSKDEEKKDKVKKDEKK 155

## RESULT 17

B81594  
hypothetical protein CP0281 [imported] - Chlamydomophila pneumoniae (strain AR39)  
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C;Accession: B81594  
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, N.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A;Reference number: A81500; MUID:20150255; PMID:10684935  
A;Accession: B81594  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-508 <REA>  
A;Cross-references: GB:AE002189; GB:AE002161; NID:g7189205; PIDN:AAF38139.1; PID:g718920  
A;Experimental source: strain AR39, HL cells  
C;Genetics:  
A;Gene: CP0281

Query Match 12.9%; Score 89.5; DB 2; Length 508;  
Best Local Similarity 24.5%; Pred. No. 11;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
QY 54 IKPVFKKIEEKEEENKPTFD-----VSKKNDPQVNHSQLNESHK 95  
DB 92 VKGVFKKTPQARPEVSPRLPSPHVQHQQLPGLGFRDRIQKRSNPEDLGMKRSYSD 151  
QY 96 EDLQREHSQKSDSTKDVTTATVLDKKNISKSTT 129  
DB 152 GDLDRVGHSDNEDSTEDSRs---EGGEFSPSKSS 182

## RESULT 18

C72074  
hypothetical protein - Chlamydomophila pneumoniae (strain CWL029)  
C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C;Accession: C72074  
R;Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A;Reference number: A72000; MUID:99206606; PMID:10192388  
A;Accession: C72074  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-508 <ARN>  
A;Cross-references: GB:AE001632; GB:AE001363; NID:g4376755; PIDN:AAD18613.1; PID:g437675  
A;Experimental source: strain CWL029  
C;Genetics:

4

Best Local Similarity 22.3%; Pred. No. 1.2e+02;  
Matches 29; Conservative 27; Mismatches 42; Indels 32; Gaps 5;

Qy 6 TIVSEEDFILPVYKGELEKGYQFDCWEISGPEGKKDAGYVINLSKDTFIKPVFKKIEBK 65  
Db 1060 TNICDENNIQJNEENSKGVISGTDN---ENKND-----MENKN 1097

Qy 66 EEENKPTFDVSKKKNPQVNH-SQLNESHKEDLQREHSQKSDTKDVTATVLDKNNIS 124  
Db 1098 DMEKKN--DMEKNDIEKNDMEKNDMEKNDMEKNDMEKNDMEKNDMEKNDMEKNDME 1148

Qy 125 SKSTNNPNK 134  
Db 1149 NKSDIENENK 1158

## RESULT 23

T33068

hypothetical protein C35E7.9 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T33068

R:Graves, T.; McDonald, R.

submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans cosmid C35E7.

A:Reference number: 221278

A:Accession: T33068

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-301 &lt;GRA&gt;

A:Cross-references: EMBL:AF067216; PIDN:AAC17524.1; GSPDB:GN00019; CESP:C35E7.9

A:Experimental source: strain Bristol N2; clone C35E7

C:Genetics:

A:Gene: CESP:C35E7.9

A:Map position: 1

A:Introns: 30/3; 193/1; 236/2

Query Match 12.7%; Score 88.5; DB 2; Length 301;  
Best Local Similarity 24.8%; Pred. No. 7.5;  
Matches 27; Conservative 17; Mismatches 38; Indels 27; Gaps 3;

Qy 25 GYQFDGWEISGFEKK-----DAGYVINLSKDTFIKPVFKI-----EE 63  
Db 10 GITTAGILAGCGGKKKGKSGSTASAAAPKADSKMKPPVENVKSKKSEKKEPKGKEEP 69

Qy 64 KKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKD 112  
Db 70 KKEEKEKSKKSEKODKK-----EAKKEDDKDEKDEKDDKD 112

## RESULT 24

A71683

hypothetical protein RP278 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000

C:Accession: A71683

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark, U.

Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630; MUID:99039499; PMID:9823893

A:Accession: A71683

A&gt;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-371 &lt;AND&gt;

A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14739.1; PID:g386083

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: RP278

Query Match 12.7%; Score 88.5; DB 2; Length 371;  
Best Local Similarity 23.1%; Pred. No. 9.5;  
Matches 34; Conservative 33; Mismatches 37; Indels 43; Gaps 9;

Qy 1 KEMSTIVSEDFILPVYKGELEKGYQFDCWEISGFEKK-DAGYVINLSKDTFIKPVFK 59  
Db 192 EQNLNTIINEE-----FRKNLEQ-----EKKIEGKKTQASEALNKK-----LKPITYK 234

Qy 60 KIEEKEE-----ENKPTFDVSKKKNPQVNH-SQLNESHKEDLQREHSQKSDTK 111  
Db 235 QMDEEREELFKLAELIPLQY-----AQANIDKHALEYAKQYOTKIENDPNYKELEKQ 286

Qy 112 DVTATVLDKNNISKSTT-----NNPN 133  
Db 287 E-----IVSKIEYKSKSKTNDIINPN 309

## RESULT 25

T20410

hypothetical protein E02A10.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T20410

R:Thomas, K.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19271

A:Accession: T20410

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-385 &lt;WIL&gt;

A:Cross-references: EMBL:Z81053; PIDN:CAB02877.1; GSPDB:GN00023; CESP:E02A10.2

A:Experimental source: clone E02A10

C:Genetics:

A:Gene: CESP:E02A10.2

A:Map position: 5

A:Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 12.7%; Score 88.5; DB 2; Length 385;  
Best Local Similarity 35.5%; Pred. No. 9.9;  
Matches 22; Conservative 14; Mismatches 19; Indels 7; Gaps 2;

Qy 59 KXIEEKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVL 118  
Db 317 KXEEKEEKEEKE--EVEKKEE-----EKKDEBPKEEKEEKEEKEEVEEKEKVE 369

Qy 119 DK 120  
Db 370 EK 371

## RESULT 26

D86432

hypothetical protein T518.14 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C:Accession: D86432

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, K.;

anssen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: D86432

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-540 &lt;STO&gt;

A:Cross-references: GB:AE005172; NID:g4587525; PIDN:AAD25756.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 12.7%; Score 88.5; DB 2; Length 540;  
Best Local Similarity 24.4%; Pred. No. 14;

Matches 31; Conservative 25; Mismatches 54; Indels 17; Gaps 4;  
QY 2 EMSSTVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVNLKDTFIKPVFKK- 60  
DB 49 EKSASFKEESDFADLKESEKK-----ALSDLKSLKEEAIVDN---TLLTKKKES 96  
QY 61 --IEEKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVL 118  
DB 97 SPMKSEKKEEVKPEAEVEKKGE--EAAEKVEEEKSEAVVTEAPKAETVEAVVTESII 154  
QY 119 DKNNTSS 125  
DB 155 PKEEVTT 161  
RESULT 27  
T47835  
hypothetical protein T209.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47835  
R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24475  
A:Accession: T47835  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <NYA>  
A:Cross-references: EMBL:AL138658  
A:Experimental source: cultivar Columbia; BAC clone T209  
C:Genetics:  
A:Map position: 3  
A:Introns: 158/2; 329/3  
A:Note: T209.90  
Query Match 12.7%; Score 88.5; DB 2; Length 644;  
Best Local Similarity 24.6%; Pred. No. 18;  
Matches 32; Conservative 18; Mismatches 53; Indels 27; Gaps 3;  
QY 9 SEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVNLKDTFIKPVFKKIEKKEE 68  
DB 534 SVADFLKRIKKNSPQKGR-----ETTSKNQKKNQGNV-----KKEND 570  
QY 69 NKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDS-----TKDVTATVLDKNNIS 124  
DB 571 HOKSDGNVKKNSKVPKRLRSSTGCKKKVEVNNNSKSSKRRKOTKETAEVATGKRGRE 630  
QY 125 SKSTNNPNK 134  
DB 631 SGKDKQPRK 640  
RESULT 28  
G88436  
protein T04A8.13 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: G88436  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/c\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: G88436  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-762 <STO>  
A:Cross-references: GB:chr\_III; PIDN:CAA84732.1; PID:g3879347; GSPDB:GN000021; CESP:T04A8  
C:Genetics:  
A:Gene: T04A8.13  
A:Map position: 3

Query Match 12.7%; Score 88.5; DB 2; Length 762;  
Best Local Similarity 25.2%; Pred. No. 21;  
Matches 32; Conservative 25; Mismatches 65; Indels 5; Gaps 3;  
QY 9 SEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVNLKDTFIKPVFKKIEKKEE 67  
DB 14 SGKQIMKPGY--DKKEGLGMDQKEIVGDDKKDKKARERKRLQDEFAB--LKDEEKDKKE 69  
QY 68 ENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKS 127  
DB 70 EAEKEKNEKEKKEKKEDGHEKKEDKEDKENENDEKKEKSKDDKKEESKEDKKEKTK 129  
QY 128 TTNNPNK 134  
DB 130 TEDNEGK 136  
RESULT 29  
T24435  
hypothetical protein T04A8.13 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T24435  
R:Palmer, S.  
submitted to the EMBL Data Library, August 1994  
A:Reference number: Z19889  
A:Accession: T24435  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-791 <WIL>  
A:Cross-references: EMBL:Z35663; PIDN:CAA84732.2; GSPDB:GN000021; CESP:T04A8.13  
A:Experimental source: clone T04A8  
C:Genetics:  
A:Gene: CESP:T04A8.13  
A:Map position: 3  
A:Introns: 31/3; 212/1; 229/3; 331/3; 406/1; 472/3; 572/1; 651/2  
Query Match 12.7%; Score 88.5; DB 2; Length 791;  
Best Local Similarity 25.2%; Pred. No. 22;  
Matches 32; Conservative 25; Mismatches 65; Indels 5; Gaps 3;  
QY 9 SEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVNLKDTFIKPVFKKIEKKEE 67  
DB 14 SGKQIMKPGY--DKKEGLGMDQKEIVGDDKKDKKARERKRLQDEFAB--LKDEEKDKKE 69  
QY 68 ENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKS 127  
DB 70 EAEKEKNEKEKKEKKEDGHEKKEDKEDKENENDEKKEKSKDDKKEESKEDKKEKTK 129  
QY 128 TTNNPNK 134  
DB 130 TEDNEGK 136  
RESULT 30  
T28676  
rhoptry protein - Plasmodium yoelii (fragment)  
C:Species: Plasmodium yoelii  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 01-Dec-2000  
C:Accession: T28676; A45521  
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
Mol. Biochem. Parasitol. 76, 329-332, 1996  
A:Title: Comparison of two members of a multigene family coding for high-molecular mass  
A:Reference number: Z20507; MUID:97077455; PMID:8920022  
A:Accession: T28676  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-2401 <STIN>  
A:Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1  
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.  
Mol. Biochem. Parasitol. 42, 241-246, 1990  
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co  
A:Reference number: A45521; MUID:91101660; PMID:2270106





A;Note: C0485w

Query March 12.6%; Score 87.5; DB 2; Length 2523;  
Best Local Similarity 32.5%; Pred.No.98;  
Matches 25; Conservative 22; Mismatches 21; Indels 9; Gaps 5;

Qy 59 KKIEKKEEENKPTFDYSKKD-NPQVNHSQLNESHREKL-QREEHSQSXDSTK--DVT 114  
|||:||::|||:|||:|||:|||:|||:|||:|||:|||:  
Db 1795 KKKIKKKDDINKKE-DINKKDIDNKONINKNDINRKDDINKKDDINKKGDIINKKDDIN 1853

Qy 115 ATVLDRKNISKSSTNN 131  
||||::|||  
Db 1854 ---KKNYNNNSNNN 1856

RESULT 35  
F71621  
hypothetical protein PF80170w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C>Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
R:Gardner, F71621  
R:Garrison, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: F71621  
A:A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-635 <GAR>  
A:CROSS-references: GB:AE001377; GB:AE001362; NID:g3845111; PIDN:AAC71823.1; PID:g3845111  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0170W

Query Match	12.5%;	Score 87;	DB 2;	Length 635;
Best Local Similarity	23.8%;	Pred. No. 23;		
Matches	30;	Conservative 20;	Mismatches 52;	Indels 24; Gaps 4;
Qy	22	LEKQYDQGWEISGFEKGKAGYVINSKDTFIKPVFKIIEEKEENKFTFDVSKKKN	81	
Db	202	LQKYNTODDDEENETIRDSKURDIYSQSKDINMSSSPNKEES	256	
Qy	82	PQVNSOLINESHRKEDLQREHSQKSDSTKDVATVILD	127	
Db	257	KDINS---DNQNKDINSSDHN-NDSTNESITTSLSNTNRNKKNRKKNININN	311	
Qy	128	TTNNPN	133	
Db	312	NNNNN	317	

RESULT 36

QRMSP1

microtubule-associated protein MAP1B - mouse

N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protein

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text\_change 01-Sep-2000

C:Accession: S07549; S44387; A33645

E:Noble, M.; Lewis, S.A.; Cowan, N.J.

J: Cell Biol. 109, 3367-3376, 1989

A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains

A:Reference number: A33645; MUID: 90094539; PMID:2480963

A:Accession: S07549

A:Molecule type: mRNA

A:Residues: 1-2464 <NO>

A:Cross-references: EMBL:X51396; NID:952999; PIDN:CAA35761.1; PID:G53000

R:Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.

Arch. Biochem. Biophys. 310, 428-432, 1994

A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.

A:Reference number: S44387; MUID: 94234720; PMID:8179328

A:Accession: S44387

A>Status: preliminary

A;Molecule type: protein  
A:Residues: 653-663,' IC'

C;Superfamily: microtubule-associated protein MAP1B  
 C;Keywords: microtubule binding; phosphoprotein; tandem repeat  
 F;589-786/Domain: microtubule Binding status experimental <MTB>  
 F;589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-699  
 R-K-E/D-X)  
 F;1861-2064/Region: 17-residue repeats  
 F;91, 116, 351, 888, 1124, 1153, 1168, 1208, 1662, 1877, 1918, 2003, 2030, 2054, 2083/Binding site: phosphate  
 F;147, 969, 1336, 1562, 1563, 1702, 1708, 1990, 2057, 2063, 22419/Binding site: phosphate (Thr) (covalent)  
 F;1953/Binding site: phosphate (Tyr) (covalent) #status predicted

```

Query Match      12.5%; Score 87; DB 1; Length 2464;
Best Local Similarity 28.4%; Pred No 1e+02;
Matches 29; Conservative 16; Mismatches 37; Indels 20; Gaps 4;

Qy 48 LSKDTFTKVPVK-KIEKGEENKPTFDVSKKDNQVNHSQLNESHKEDLRE-----101
   :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 632 VTKDQVVVKIEKTLBEKKEE--KPKKEVVKDKTFL--KKDKPRKEEVKKEIKKEI 686

Qy 102 -----EHSOKSDSTKDVTATVLDDKNNISKSTTNNPK 134

Db 687 KGBERKELKEVKETPLKDAKEVVKKEKKEKVKKEKEPKK 728

```

RESULT 37

T28391  
ORF MSV230 hypothetical protein - Melanoplus sanguinipes entomopoxvirus  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T28391  
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28391  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-670 <AFO>  
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97713.1; PID:g4049753  
C:Genetics:  
A:Note: MSV230

Query Match	12.4%	Score	86.5	DB	2	Length	670
Best Local Similarity	22.7%	Pred. No.	26				
Matches	35	Conservative	33	Mismatches	65	Indels	21
Gaps	5						
Qy	1	KEMSSSTI-----	VSEEDFTLPVYKGELEKGY-----	QPDGWEISGFEGKKDA	42		
Db	39	KQIISTLLKFNFDKTEMCGVSEKVFQLINNKSA	SEKYSVDVSSIDSSDS	98			
Qy	43	GYVINLSKDTFKIPFKKIE-EKKEEENKPTFFVDSKKKNPQVNH	SOL-NESHKEDLQR	100			
Db	99	GVNIDESQNSDSKVNINKLENSQNSDSKVNIDESQNSDS	-KVNINKLENSQNSDSKVN	157			
Qy	101	EEHSQKSDSKDTKVATVLDKNNISSKSTTTNNPNK	134				
Db	158	IDPSQNSDSKVNIDESQNSDSKVNIDESQNSDSK	191				

RESULT 38

T49989  
hypothetical protein F12B17.150 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49989  
R:Bayan, M.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: 225026  
A:Accession: T49989  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-792 <BEV>

RESULT 40  
T19006  
ankyrin related protein C06C3.1 - *Caenorhabditis elegans*  
N:Contains: myosin-light-chain-phosphatase (EC 3.1.3.53)  
C:Species: *Caenorhabditis elegans*  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Mar-2000  
C/Accession: T19006; T22086  
R/Berks, M.  
submitted to the EMBL Data Library, August 1994  
A:Reference number: Z19058

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 6.22948 Seconds  
(without alignments)  
1011.574 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTIVSEDFILPYKG.....ATVLDKNISSKSTTNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97.5	14.0	1345	1 YH00 YEAST	P38800 saccharomyc
2	95	13.6	348	1 CYL2 HUMAN	Q14093 homo sapien
3	93	13.4	1398	1 TOP2 PLAFK	P41001 plasmodium
4	92.5	13.3	2468	1 MAPB HUMAN	P46821 homo sapien
5	89	12.8	1888	1 YDT2 SCHPO	Q14207 schizosacch
6	87.5	12.6	1702	1 IGA2 HAEIN	P45384 haemophilus
7	87	12.5	2464	1 MAPB MOUSE	P14873 mus musculu
8	86	12.4	443	1 GLNA PYRKO	Q08467 pyrococcus
9	85	12.2	720	1 IF2 STAEF	P23374 ascbolus i
10	85	12.2	1202	1 DPOM ASCIM	Q8cst4 staphylococ
11	84.5	12.1	1332	1 SPT7 YEAST	P35177 saccharomyc
12	84.5	12.1	1694	1 IGA0 HAEIN	P44969 haemophilus
13	84	12.0	778	1 YFK8 YEAST	P3610 saccharomyc
14	83.5	12.0	479	1 UR21 HUMAN	Q15695 homo sapien
15	83	11.9	258	1 IF31 HUMAN	O75822 homo sapien
16	83	11.9	270	1 TONB HAEIN	P42872 haemophilus
17	82.5	11.9	439	1 GLNA PYRAB	Q8uy99 pyrococcus
18	82	11.8	279	1 YNB4 YEAST	P49957 saccharomyc
19	82	11.8	427	1 YKT5 YEAST	P36046 saccharomyc
20	82	11.8	439	1 GLNA PYRFU	Q05907 pyrococcus
21	82	11.8	439	1 GLNA PYRWO	P36687 pyrococcus
22	82	11.8	949	1 IF2 HELVI	Q9zm46 helicobacte
23	81.5	11.7	572	1 LMD1 HUMAN	P29536 homo sapien
24	81.5	11.7	752	1 DRS1 YEAST	P32892 saccharomyc
25	81.5	11.7	2459	1 MAPB RAT	P15205 rattus norv
26	81	11.6	797	1 PAT1 YEAST	P25644 saccharomyc
27	81	11.6	853	1 YCG1 YEAST	P25588 saccharomyc
28	81	11.6	1325	1 YAE6 SCHPO	Q09847 schizosacch
29	80.5	11.6	256	1 TONB PASMO	Q9zn79 pasteurella
30	80.5	11.6	793	1 YF06 MYCFN	P37280 mycoplasma
31	80.5	11.6	914	1 PPFA BACSU	P39793 bacillus su
32	80.5	11.6	1006	1 RAT1 YEAST	Q02792 saccharomyc
33	80	11.5	688	1 LTP STAEF	Q02510 staphylococ

#### ALIGNMENTS

##### RESULT 1

YH00 YEAST 80 11.5 1220 1 IF2P HUMAN  
 ID YH00 YEAST STANDARD; PRT; 1345 AA.  
 AC P38800;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Hypothetical 149.7 kDa protein in IRE1-KSP1 intergenic region.  
 GN YHR080C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP  
 RC SEQUENCE FROM N.A.  
 RX STRAIN=S288C / AB972;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
 Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
 Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
 Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,  
 Nham M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,  
 Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
 Vaudin M.;  
 RA "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
 VIII.";  
 RL Science 265:2077-2082(1994).  
 CC -1- SIMILARITY: TO YEAST YFL042C.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U10556; AAB68895.1; ;  
 DR PIR; S46817; S46817.  
 DR SGD; S0001122; YHR080C.  
 DR InterPro; IPR004182; GRAM\_dom.  
 DR Pfam; PF02893; GRAM; 1.  
 DR SMART; SM00568; GRAM; 1.  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 1198 1218 POTENTIAL.  
 SQ SEQUENCE 1345 AA; 149679 MW; 2FDAB9A686564C2 CRC64;

Query Match 14.0%; Score 97.5; DB 1; Length 1345;

Best Local Similarity 27.0%; Pred. No. 4.3;

Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

Qy 19 KGELEKYGQDFGWEISCFEGKK-DAGYVNLKSDTFIKPVFKIEEKEENKTFDYSK 77

Db 1109 KGAEIKG-----SVGEQKVSVDYMLSELRLDII-----SRAKSKKPVKVMK 1149

Qy 78 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLD--KNNISSKSTTN---NP 132



Db 1199 EFQNVLSNREESNK--FKVARKQ-----GPSMKKKKKKKLSSDESEGGDTSDSSE 1251  
QY 112 DVTATVLDKKNISKSSTNNPN 133  
Db 1252 FLVNTLNIKKNTKTKTSSNN 1273  
RESULT 4  
MAPB HUMAN  
ID MAPB\_HUMAN STANDARD; PRT; 2468 AA.  
AC P46821; 1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain  
LC1].  
OS Homo sapiens (Human).  
GN MAP1B.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal brain;  
RX MEDLINE=95104835; PubMed=7806212;  
RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;  
RT "Cloning of human microtubule-associated protein 1B and the  
RT identification of a related gene on chromosome 15.";  
RL Genomics 22:273-280(1994).  
CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS  
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
CC STABILIZING MICROTUBULES.  
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
CC WITH MAP1A AND MAP1B PROTEINS.  
CC -!- DOMAIN: Has a highly basic region with many copies of the sequence  
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
CC responsible for the binding of MAP1B to microtubules.  
CC -!- PM: LC1 is coexpressed with MAP1B. It is a polypeptide generated  
CC from MAP1B by proteolytic processing. It is free to associate with  
CC both MAP1A and MAP1B. It interacts with the amino-terminal region  
CC of MAP1B (By similarity).  
CC -!- SIMILARITY: TO MAP1A.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L06237; AAA18904.1; -  
DR DR  
DR MIN; 157129; -  
DR GO; GO:0005875; C:Microtubule associated complex; TAS.  
DR InterPro; IPR000102; MAP1B neuraxin.  
DR Pfam; PF00414; MAP1B neuraxin; 10.  
DR PROSITE; PS00230; MAP1B NEURAXIN; 6.  
KW Microtubules; Repeat; Phosphorylation.  
FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.  
FT REPEAT 1878 1894 MAP1B 1.  
FT REPEAT 1895 1911 MAP1B 2.  
FT REPEAT 1912 1928 MAP1B 3.  
FT REPEAT 1929 1945 MAP1B 4.  
FT REPEAT 1946 1962 MAP1B 5.  
FT REPEAT 1963 1979 MAP1B 6.  
FT REPEAT 1997 2013 MAP1B 7.  
FT REPEAT 2014 2030 MAP1B 8.  
FT REPEAT 2031 2047 MAP1B 9.  
FT REPEAT 2048 2064 MAP1B 10.

FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
FT KKEE AND KKEI/V REPEATS).  
SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;  
Query Match 13.3%; Score 92.5; DB 1; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 19;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
QY 48 LSKDTFIKVPFKKIEEKEENKPTFDVSKKKONPQVNHSQLSHREKDLQRE-----E 102  
Db 638 VKKETVKP-----EDKKEEKEKKEVAKEDKTPI---KKEEKPKEEKKVKEIK 689  
QY 103 HSOKSDSTKDV 113  
Db 690 KEEKKEPKKEV 700  
RESULT 5  
YDT2 SCHPO  
ID YDT2 SCHPO STANDARD; PRT; 1888 AA.  
AC O14207;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein C6B12.02c in chromosome 1.  
GN SPAC6B12.02C.  
OS Schizosaccharomyces pombe (Fission Yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwallam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehtach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Dague R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,  
RA Cerrutti L., Lowe I., McCombie W.R., Paulsen I., Potashkin J.,  
RA Sipakowski G.V., Ussery D., Barrel B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z98531; CAB11064.1; -  
DR DR  
DR PIR; T39009; T39009.





CC -!- SIMILARITY: TO MAP1A. -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)). -----  
CC  
CC EMBL; X51396; CRA35761.1; -.  
CC PIR; S07549; CRMSP1.  
CC MGD; MGI:1306778; Mtap1b.  
CC GO; GO:0016358; P:dendrite morphogenesis; IMP.  
CC GO; GO:0001578; P:microtubule bundling; IMP.  
CC InterPro: IPR001012; MAP1B neuraxin.  
CC Pfam; PF00414; MAP1B neuraxin; 10.  
CC PROSITE; PS00230; MAP1B NEURAXIN; 7.  
CC Microtubules; Repeat; Phosphorylation.  
CC CHAIN ? 2464 MAP1 LIGHT CHAIN LCL1.  
FT REPEAT 1874 1890 MAP1B 1.  
FT REPEAT 1891 1907 MAP1B 2.  
FT REPEAT 1908 1924 MAP1B 3.  
FT REPEAT 1925 1941 MAP1B 4.  
FT REPEAT 1942 1958 MAP1B 5.  
FT REPEAT 1959 1975 MAP1B 6.  
FT REPEAT 1993 2009 MAP1B 7.  
FT REPEAT 2010 2026 MAP1B 8.  
FT REPEAT 2027 2043 MAP1B 9.  
FT REPEAT 2044 2060 MAP1B 10.  
FT DOMAIN 589 787 KKEE AND KKEI/V REPEATS).  
SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDBDA87 CRC64;  
  
Query Match 12.5%; Score 87; DB 1; Length 2464;  
Best Local Similarity 28.4%; Pred. No. 51;  
Matches 29; Conservative 16; Mismatches 37; Indels 20; Gaps 4;  
  
Qy 48 LSKDTFTKPVFK-KIEEKEEENKPTFDVSKKDKNPVNHSQLNESHKEDLORE----- 101  
Db 632 VTOKVVKKEIKTKLEEKKEE-KPKKEVVKKEKDTPL---KDEKPKKEVKKEIKKEI 686  
Qy 102 -----EHSQKSDSTQKVATVTLVDKNISSKSTNNPNK 134  
Db 687 KKEERKELKKEVKKEKPTLKDAAKKEVKKEEKEKKEPKK 728  
  
RESULT 8  
GLNA\_PVRKO STANDARD; PRT; 443 AA.  
AC O08467;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
GN GLNA.  
OS Pyrococcus kodakaraensis.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Thermococcus.  
OX NCBI\_TaxID=69014;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97316461; PubMed=9172372;  
RA Adul Rahman R.N.Z., Jongsareejit B., Fujiwara S., Imanaka T.;  
RT "Characterization of recombinant glutamine synthetase from the  
RT hyperthermophilic archaeon Pyrococcus sp. strain KOD1.";  
RL Appl. Environ. Microbiol. 63:2472-2476(1997).  
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
CC L-glutamine.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY. -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)). -----  
CC  
CC EMBL; D86222; BAA20530.1; -.  
CC HSP; P06201; ILGR.  
CC InterPro; IPR001691; GLN synth.  
CC InterPro; IPR004809; GlnA\_adenyln.  
CC Pfam; PF00120; gln-synt\_N; 1.  
CC Pfam; PF03951; gln-synt\_C; 1.  
CC ProDom; PD001057; Gln\_synt\_C; 1.  
CC TIGRFAMs; TIGR00653; GlnA\_1.  
CC PROSITE; PS00180; GlnA\_1; 1.  
CC PROSITE; PS00181; GlnA\_ATP; 1.  
KW Ligase.  
FT BINDING 362 362 AMP (UNDER CONDITIONS OF ABUNDANT  
FT BINDING 443 AA; 50259 MW; 9426DCCFEF18168 CRC64;  
SQ SEQUENCE 443 AA; 50259 MW; 9426DCCFEF18168 CRC64;  
  
Query Match 12.4%; Score 86; DB 1; Length 443;  
Best Local Similarity 17.8%; Pred. No. 10;  
Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;  
  
Qy 18 YKGELEKGYQFDGWEISGPEGKDGAGVINLSKDTFI----- 54  
Db 42 YEEAEDGVSGFDGSSIPGFEIGSDLIKADPSTYAEIPWEGIGRVGYIKGDEPYQA 101  
Qy 55 --KPVFKKIEEKKKEEN-----KPTFDVSKKDK----- 80  
Db 102 DPRGILKRVLERLEKEGLKAHIGPEFEYFKNGTWELHIPDSGGYFDLVGLDKAREIR 161  
Qy 81 -----NPQVNHSQLNESHKEDLOREHSQKSD---STKDVATVLD 119  
Db 162 REIATLYMPYLGLKPEVLHHEVGKAQHEIDFRYDEALRTADNIVSFKHVRAVAE 215  
  
RESULT 9  
IF2\_STAEP STANDARD; PRT; 720 AA.  
AC Q8CST4;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Translation initiation factor IF-2.  
GN INFB OR SE0945.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
RA Chen Z., Wen Y.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: One of the essential components for the initiation of  
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous  
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.  
CC Also involved in the hydrolysis of GTP during the formation of the  
CC 70S ribosomal complex (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY. -----  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)). -----

```
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE016747; AA04542.1; -.
DR HAMAP; MF_00100; -.
DR PFam; PF00009; GTP_EFTU; 1.
DR PFam; PF03144; GTP_EFTU_D2; 1.
DR PFam; PF04760; IF2_N; 2.
DR PRODom; PD186100; IF2_N; 1.
DR TIGSFAMs; TIGR00487; IF-2; 1.
DR TIGSFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 225 373 G-DOMAIN.
FT NP_BIND 231 238 GTP (BY SIMILARITY).
FT NP_BIND 277 281 GTP (BY SIMILARITY).
FT NP_BIND 331 334 GTP (BY SIMILARITY).
SQ SEQUENCE 720 AA; 79343 MW; 07PB5A6A59CF970C CRC64;

Query Match 12.2%; Score 85; DB 1; Length 720;
Best Local Similarity 28.0%; Pred. No. 20;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 61 IEEKKEENKPTFDVSKKD-----NPQVNHSQLNESHKEDLQREHSQKSDSTK----D 112
DB 38 LEEQIKALDKKFKASQAKDTNKQNTQNNHQSKNNKQNSDKKQSKNNKPTKKKEQN 97
QY 113 VTATVLDKNNISKSTTNPNK 134
DB 98 NKGKQNNKNNKNNKQNNK 119

RESULT 10
DPOM_ASCIM
ID DPOM_ASCIM STANDARD; PRT; 1202 AA.
AC P22374;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable DNA polymerase (EC 2.7.7.7).
OS Ascobolus immersus.
OG Mitochondrion.
OG Plasmid pA12.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;
OC Pezizales; Ascobolaceae; Ascobolus.
OX NCBI_TaxID=5191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2/1;
RX MEDLINE=90066356; PubMed=2573821;
RA Kempken F., Meinhardt F., Esser K.;
RT "In organello replication and viral affinity of linear,
RT extrachromosomal DNA of the ascomycete Ascobolus immersus.";
RL Mol. Gen. Genet. 218:523-530 (1989).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC -!- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG. TO
CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X15982; CAA34106.1; --
DR PIR; S05362; S05362.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR004868; DNA_pol_B_2.
DR -----

DR PFam; PF03175; DNA_pol_B_2; 1.
DR SMART; SM00486; POLBc; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Plasmid; Mitochondrion.
SQ SEQUENCE 1202 AA; 138279 MW; 51D41FCEBDBF2CDE CRC64;

Query Match 12.2%; Score 85; DB 1; Length 1202;
Best Local Similarity 25.6%; Pred. No. 34;
Matches 33; Conservative 21; Mismatches 55; Indels 20; Gaps 5;

QY 19 KGELEKGYQFDGWEISGEGKKDAGYVNLISKDFIKPVPFKKIE-----EKKEEENK-P 71
DB 372 KGSU-----SFDPKPLKTEGTGYANYTFPIKDIVKDKINKINPGLDLPTWDLKWP 427
QY 72 TFDVSKKNDPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNI-----S 124
DB 428 NLKLNKDKTSGEIRMTIKNNQSYDI--IGHMIINDGENVITFNRAVDNSIIKIFTVTD 485
QY 125 SKSTTNPN 133
DB 486 SMGNTNDPN 494

RESULT 11
SEPT_YEAST
ID SEPT_YEAST STANDARD; PRT; 1332 AA.
AC P35177;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Transcriptional activator SP17.
GN SP17 OR YBR081C OR YBR0739.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95229044; PubMed=7713415;
RA Gansheroff L.J., Dollard C., Tan P., Winston F.;
RT "The Saccharomyces cerevisiae SP17 gene encodes a very acidic protein
RT important for transcription in vivo.";
RL Genetics 139:523-536 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95076715; PubMed=7985423;
RA van der Aart Q.J.M., Barthe C., Doignon F., Aigle M., Crouzet M.,
RA Steensma H.Y.;
RT "Sequence analysis of a 31 kb DNA fragment from the right arm of
RT Saccharomyces cerevisiae chromosome II.";
RL Yeast 10:959-964 (1994).
RN [3]
RP SEQUENCE OF 1-835 FROM N.A.
RC STRAIN=S288C;
RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
RA Vissers S.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 463-523 FROM N.A.
RX MEDLINE=92285152; PubMed=1350857;
RA Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J.,
RA David I.B.;
RT "The bromodomain: a conserved sequence found in human, Drosophila and
RT yeast proteins.";
RL Nucleic Acids Res. 20:2603-2603 (1992).
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY
CC OTHER GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -----
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L22537; AAC37424.1; -;  
 EMBL; X76294; CRA53940.1; -;  
 EMBL; Z35950; CRA85026.1; -;  
 EMBL; M87651; AAA35087.1; -;  
 PIR; S41552; S41552.  
 HSSP; Q92831; 1B91.  
 TRANSFAC; T04835; -;  
 SGD; S0000285; SPT7.  
 GO; GO:0000124; C:SAGA complex; IDA.  
 InterPro; IPR001487; Bromodomain.  
 Pfam; PF00439; bromodomain; 1.  
 PRINTS; PR00503; BROMODOMAIN.  
 SMART; SM00297; BROMO; 1.  
 PROSITE; PS00633; BROMODOMAIN\_1; 1.  
 PROSITE; PS0014; BROMODOMAIN\_2; 1.  
 Transcription regulation; Nuclear protein; Activator; Bromodomain.  
 KW DOMAIN 458 528 BROMODOMAIN.  
 FT SEQUENCE 1332 AA; 152616 MW; 083B63624669244F CRC64;

Query Match 12.1%; Score 84.5; DB 1; Length 1332;  
 Best Local Similarity 24.3%; Pred. No. 41;  
 Matches 36; Conservative 24; Mismatches 61; Indels 27; Gaps 6;

QY 6 TIVSEEDFLPVYKGELEKGYQFDGWEISGPEGKKDAGYVNLKSDTFIK---PVFKKIE 62  
 DB 550 TIRRADLEKIEDEKDKDYELDEEEVAGSRKG-----LNGAHMLAKENGKVKSEKDS 605  
 QY 63 EKKEEENKPTFD-----VSKKKD-----NPQVNHSQLNESHKEDLQR-EEH 103  
 DB 606 SKTVKDEAPTNDKLTSTVPEGEKEKDTASSTVTVHENVNKNKENGKNEEQDMVEES 665  
 QY 104 SKQSDSTKDVATVLD-KNVLSSKSTTN 130  
 DB 666 SKTEDSSKDAAKKDTEDGLQDKTAEN 693

RESULT 12  
 ID IGA0 HAEIN STANDARD; PRT; 1694 AA.  
 AC P44969;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAI protease).  
 GN IGA OR IGAI OR HI0990.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OK NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Serotype D;  
 RA Wright A., Fishman Y., Tai F., Plaut A.G.;  
 RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA McLavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.P., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.;

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RL Science 269:496-512(1995).  
 CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
 CC PRODUCING INTACT FC AND FAB FRAGMENTS.  
 CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
 CC certain Pro-|-Xaa bonds in the hinge region. No small molecule  
 CC substrates are known.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X59800; -; NOT ANNOTATED\_CDS.  
 EMBL; U32779; AAC22651.1; -;  
 PIR; H64106; H64106.  
 MEROPS; S06.001; -.  
 TIGR; HI0990; -.  
 InterPro; IPR006315; Autotransport.  
 InterPro; IPR005546; Autotransporter.  
 InterPro; IPR000710; IGA\_S6.  
 InterPro; IPR004899; Pertactin.  
 Pfam; PF03797; Autotransporter; 1.  
 Pfam; PF02395; IGAI; 1.  
 Pfam; PF03212; Pertactin; 1.  
 PRINTS; PR00321; IGASERPTASE.  
 TIGRFAMs; TIGR01414; autotrans\_barl; 1.  
 DR Hydrolyase; Serine protease; transmembrane; Zymogen; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 25  
 CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.  
 FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).  
 FT ACT\_SITE 288 288 PROBABLE.  
 FT CONFLICT 253 254 EN -> GV (IN REF. 1).  
 FT CONFLICT 272 272 G -> A (IN REF. 1).  
 FT CONFLICT 464 464 G -> E (IN REF. 1).  
 FT CONFLICT 866 866 S -> T (IN REF. 1).  
 FT CONFLICT 1036 1036 A -> D (IN REF. 1).  
 FT CONFLICT 1074 1074 A -> G (IN REF. 1).  
 FT CONFLICT 1421 1421 A -> G (IN REF. 1).  
 FT CONFLICT 1545 1545 H -> T (IN REF. 1).  
 SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;

Query Match 12.1%; Score 84.5; DB 1; Length 1694;  
 Best Local Similarity 26.1%; Pred. No. 53;  
 Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

QY 46 INLSKDTFKIPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHS- 104  
 DB 1288 INTGSAITETAESDKPQETAASTEDTADNSKSKPNR 1379  
 QY 105 --QKSDSTKDVATVLDKNVLSSKSTTNPNK 134  
 DB 1348 SQPQETSAETAASTEDTADNSKSKPNR 1379  
 RESULT 13  
 YFK8 YEAST  
 ID YFK8 YEAST STANDARD; PRT; 778 AA.  
 AC P43610;



```

DR ENBL; BC002719; AA002719.1; -.
DR Genew; HGNC:3270; EIF3S1.
DR GK; O75822; -.
DR MTM; 603910; -.
DR DR GO; GO:0005852; C:eukaryotic translation initiation factor 3 . . .; TAS.
DR DR GO; GO:0003743; F:translation initiation factor activity; TAS.
DR DR GO; GO:0006446; P:regulation of translational initiation; TAS.
DR KW Initiation factor; Protein biosynthesis.
FT DOMAIN 2 8 POLY-ALA.
FT DOMAIN 29 32 POLY-GLY.
FT DOMAIN 53 57 POLY-ASP.
FT DOMAIN 218 224 POLY-LYS.
FT CONFLICT 40 40 E -> G (IN REF. 1).
FT CONFLICT 141 141 A -> T (IN REF. 2).
SQ SEQUENCE 258 AA; 29062 MW; 8362423542445AA CRC64;

Query Match 11.9%; Score 83; DB 1; Length 258;
Best Local Similarity 25.4%; Pred. No. 9.6;
Matches 30; Conservative 25; Mismatches 51; Indels 12; Gaps 4;

QY 18 YKGELEKGYQFGWETSGFEGKKDAGY--VINLSKDTFTKPVFKIEEKKEENKPTFDV 75
Db 39 WEGEDDEDVKNWDDDDDEKKEAEVKEVISEK---KKIAEKIKERQKKRQERI 95
QY 76 SKKKNPQ-----VNHSQLNESHKEDIQREHHSQKSDTKDVTATV--LDKNNISK 126
Db 96 KRLLEPEBPVKVLTPEQLDGLRLKLLQFESLELAKETFGVNNVAYGIDAMNPSR 153

RESULT 16
TONB_HABIN STANDARD; PRT; 270 AA.
ID TONB_HABIN STANDARD; PRT; 270 AA.
AC P42872;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE TonB protein.
GN TONB OR HI0251.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTHI TNI06;
RX MEDLINE=9424537; PubMed=8188372;
RA Jarosik G.P., Sanders J.D., Cope L.D., Muller-Eberhard U.,
RA Hansen E.J.;
RT "A functional tonB gene is required for both utilization of heme and
RL virulence expression by Haemophilus influenzae type b.";
RL Infect. Immun. 62:2470-2477(1994).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.P., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Frichman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RP "whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER

```

MEMBRANE PROTEINS. REQUIRED FOR HEME UTILIZATION AND VIRULENCE.  
-!- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC  
MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE  
PERIPLASM (BY SIMILARITY).  
-!- SIMILARITY: BELONGS TO THE TONB FAMILY.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; U04996; AAA60460.1; --  
DR EMBL; U32711; AAC21917.1; --  
DR PIR; F64057; F64057.  
DR TIGR; HI0251; --  
DR InterPro; IPR003538; TonB.  
DR InterPro; IPR006260; TonB\_C.  
DR Pfam; PF03544; TonB; 1.  
DR PRINTS; PR01374; TONBPROTEIN.  
DR TIGRFAMs; TIGR01352; tonB Cterm; 1.  
KW Transport; Protein transport; Inner membrane; Periplasmic;  
Transmembrane; Signal-anchor; Repeat; Virulence; Complete proteome.  
FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 8 27 SIGNAL-ANCHOR (POTENTIAL).  
FT DOMAIN 28 270 PERIPLASMIC (POTENTIAL).  
FT DOMAIN 70 77 4 X 2 AA TANDEM REPEATS OF E-P.  
FT DOMAIN 58 90 GLU/PRO-RICH.  
FT DOMAIN 98 143 GLU/LYS/PRO-RICH.  
FT VARIANT 17 17 A -> V (IN STRAIN TN106).  
FT VARIANT 31 31 E -> K (IN STRAIN TN106).  
FT VARIANT 66 66 D -> N (IN STRAIN TN106).  
FT VARIANT 76 84 MISSING (IN STRAIN TN106).  
FT VARIANT 116 116 G -> E (IN STRAIN TN106).  
FT VARIANT 120 120 G -> E (IN STRAIN TN106).  
FT VARIANT 134 134 K -> Q (IN STRAIN TN106).  
FT VARIANT 140 140 E -> D (IN STRAIN TN106).  
FT VARIANT 146 146 E -> K (IN STRAIN TN106).  
FT VARIANT 226 226 K -> R (IN STRAIN TN106).  
SQ SEQUENCE 270 AA; 29193 MW; 4CDA54B46F7D10A0 CRC64;  
Query Match 11.9%; Score 83; DB 1; Length 270;  
Best Local Similarity 32.1%; Pred. No. 10;  
Matches 26; Conservative 16; Mismatches 33; Indels 6; Gaps 3;  
QY 50 KDTFTKVPKKEEKEENKFTFDVSKKKNPOVNHSQLNESHKEDLQREHSQKSDS 109  
DB 93 EDPTKPEPKKEPEKEPKPK---GKPKGPK-NKPKVKPKPKINKE--LPKGD 146  
QY 110 TKDVTATVLDKNNISSKSTTN 130  
DB 147 NIDSSANVNDKASTTSAANSN 167  
RESULT 17  
GLNA PYRAB STANDARD; PRT; 439 AA.  
AC Q9UY39;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
GN GLNA OR PYRAB16090 OR PAB1292.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID:29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GE5 / Orsay;  
RX PubMed=12622808;

Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,  
Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,  
Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;  
"An integrated analysis of the genome of the hyperthermophilic  
archaeon Pyrococcus abyssi.";  
Mol. Microbiol. 47:1495-1512(2003).  
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
L-glutamine.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; AJ248298; CAB50513.1; --  
DR PIR; C75009; C75009.  
DR HSSP; P06201; ILGR.  
DR InterPro; IPR001691; GLN synth.  
DR InterPro; IPR004809; GlnA.  
DR InterPro; IPR001637; GlnA adenyltn.  
DR Pfam; PF00120; gln-synt; 1.  
DR Pfam; PF03951; gln-synt\_N; 1.  
DR ProDom; PD001057; Gln\_synt\_C; 1.  
DR TIGRFAMs; TIGR00653; GlnA; 1.  
DR PROSITE; PS00180; GLNA\_1; 1.  
DR PROSITE; PS00181; GLNA\_ATP; 1.  
KW Ligase; Complete proteome.  
FT BINDING 358 358 AMP (UNDER CONDITIONS OF ABUNDANT  
GLUTAMINE) (BY SIMILARITY).  
FT SEQUENCE 439 AA; 49742 MW; 62CCFD3970A98AF0 CRC64;  
Query Match 11.9%; Score 82.5; DB 1; Length 439;  
Best Local Similarity 17.8%; Pred. No. 18;  
Matches 28; Conservative 24; Mismatches 36; Indels 69; Gaps 4;  
QY 22 LEKGYQFDGWEISGPEGKKDAGYVINLSKDTFI-----KP----- 56  
DB 42 IEDGISFDGSSVPGFGIEDSLIFKADPDYVEVPDNNVARVYGYIKDGKPYGADPRG 101  
QY 57 VFKEIEKEEEN-----KPTFDVSKKKN----- 81  
DB 102 VLKRVIEKLAEIMGKAYIGPEPEFYLFKNGSWELEIPDVGGYFDILTLDKAKDIKREIA 161  
QY 82 -----POVNHSQLNESHKEDLQREHSQKSDS 109  
DB 162 EYMFSGVLVPELHHEHVGKQAEIDFYDEALKTADN 198  
RESULT 18  
YME4 YEAST  
ID YME4 YEAST STANDARD; PRT; 279 AA.  
AC P49957;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Hypothetical 32.4 kDa protein in TAF40-ERV25 intergenic region.  
GN YML014W OR YN9571.04.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RX PubMed=9169872;  
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,

RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RL XIII,"  
 CC Nature 387:90-93 (1997).  
 CC -1- SIMILARITY: TO S.POMBE SPAC13D6.03C AND SOME, TO C.ELEGANS  
 CC C14B1.5.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z49810; CA89938.1; -  
 CC PIR; S55105; S55105.  
 CC SGD; S0004476; YML014W.  
 CC InterPro; IPR001601; Methyltransf.  
 CC InterPro; IPR000051; SAM\_bind.  
 CC Hypothetical protein.  
 CC SEQUENCE 279 AA; 32438 MW; 2F8795BB3C640D3B CRC64;  
 SQ  
 Query Match 11.8%; Score 82; DB 1; Length 279;  
 Best Local Similarity 28.9%; Pred. No. 12;  
 Matches 24; Conservative 14; Mismatches 39; Indels 6; Gaps 3;  
 YQ 31 WEISGFEGKDGAGVNLNLSKDTFTKPVFKIEEKEENKPTFDVSKKONPQVNHSQLN 90  
 DB 145 WALE--QCSSRRRGHEGQDFVFWLPSKSKPKTKSTPP---AKVTRPKPLNMLNP 199  
 YQ 91 ESHRKEDLQR-EHESQKSDSTKD 112  
 DB 200 PKERSEVLQRWKEQQRKSLDD 222  
 RESULT 19  
 ID\_YKTS\_YEAST STANDARD; PRT; 427 AA.  
 AC P36046;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE Hypothetical 47.4 kDa protein in PAS1-MST1 intergenic region.  
 GN YKL195W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,  
 RA Guerreiro P., Rodrigues-Pousada C.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z28195; CA82039.1; -  
 CC PIR; S38032; S38032.  
 CC SGD; S0001676; YKL195W.  
 CC Hypothetical protein.  
 CC SEQUENCE 427 AA; 47416 MW; 44386D250DE5DEDA CRC64;  
 SQ  
 Query Match 11.8%; Score 82; DB 1; Length 427;  
 Best Local Similarity 27.4%; Pred. No. 19;  
 Matches 29; Conservative 18; Mismatches 43; Indels 16; Gaps 3;

QY 21 ELEKGYQDFGWEISGFEGKDGAGVNLNLSKDTFTKPVFKIEEKEENKPTFDVSKKOD 80  
 DB 142 ETEAGPOLGDKTGASKVAEDGELWLAEDN-----KSEDKDTDESK-----VSTKOD 191  
 QY 81 NPQVNHSQLNESHKEDLQREHSQKSDSTKDTATVLDKNNISSK 126  
 DB 192 -----EQSNEDNATANNQDENISSENSENSTKTLDDNNAGSSE 231  
 RESULT 20  
 ID\_GLNA\_PYRFU STANDARD; PRT; 439 AA.  
 AC Q05907;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
 GN GLNA OR PF0450.  
 OS Pyrococcus furiosus.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=2261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
 RX MEDLINE=94365840; PubMed=7916055;  
 RA Brown J.R., Masuchi Y., Robb F.T., Doolittle W.F.;  
 RT "Evolutionary relationships of bacterial and archaeal glutamine  
 RT synthetase genes,"  
 RL J. Mol. Evol. 38:566-576 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
 RT "The complete sequence of the Pyrococcus furiosus genome,"  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
 CC L-glutamine.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L12410; AAA71968.1; -  
 CC EMBL; AE010168; AAL80574.1; -  
 CC HSP; P08201; ILGR.  
 CC InterPro; IPR001691; GLN\_synth.  
 CC InterPro; IPR004809; GLNA.  
 CC InterPro; IPR001637; GLNA\_adenyltn.  
 CC Pfam; PF00120; gln-synt\_1.  
 CC Pfam; PF03951; gln-synt\_N; 1.  
 CC ProDom; PD001057; gln\_synt\_C; 1.  
 CC TIGRFAMs; TIGR00653; GLNA; 1.  
 CC PROSITE; PS00180; GLNA\_1; 1.  
 CC PROSITE; PS00181; GLNA\_ATP; 1.  
 CC Ligase; Complete proteome.  
 KW BINDING 358 358  
 FT AMP (UNDER CONDITIONS OF ABUNDANT  
 FT GLUTAMINE) (BY SIMILARITY).  
 FT K -> G (IN REF. 1).  
 FT M -> L (IN REF. 1).  
 FT A -> S (IN REF. 1).  
 FT I -> L (IN REF. 1).  
 FT S -> N (IN REF. 1).  
 FT E -> G (IN REF. 1).  
 FT S -> S (IN REF. 1).  
 FT MPKDTK -> IPPDTE (IN REF. 1).  
 FT SEQUENCE 439 AA; 50183 MW; 96563A8B9A0E0892 CRC64;

Query Match 11.8%; Score 82; DB 1; Length 439;  
 Best Local Similarity 18.1%; Pred. No. 20;  
 Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

QY 25 GYQFDGWEISGEGKKGAGYVNLKDTFI-----KP-----VFK 59  
 DB 45 GLSFGSSVPGFQIGEDSLVFKADPDYVEVPWNVARVGYFYKDKPYGADPRGILK 104  
 QY 60 KIEEKEEBEN-----KPTFDVSKKKDN-----81  
 DB 105 RALELEKEGKYKAYIGPEPEFYLFKNGTWLEIPDVGGYFDILTLDKARDIRREIAEYM 164  
 QY 82 -----PQVNSQLNESHRKDLQREHSQKSD---STKDVATVLDKNNI 123  
 DB 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKTADNIVSFKYITKAVAEHGL 215

RESULT 21  
 ID GLNA PYRMO STANDARD; PRT; 439 AA.  
 AC P36687;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glutamine synthetase I (EC 6.3.1.2) (Glutamate--ammonia ligase) (GSI).  
 GN GLNA.  
 OS Pyrococcus woesei.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=2262;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93259940; PubMed=8098326;  
 RA Tiboni O., Cammarano P., Sanangelantoni A.M.;  
 RT "Cloning and sequencing of the gene encoding glutamine synthetase I  
 from the archaeum Pyrococcus woesei: anomalous phylogenies inferred  
 from analysis of archaeal and bacterial glutamine synthetase I  
 sequences.";  
 RL J. Bacteriol. 175:2961-2969(1993).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
 L-glutamine.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X60161; CAA42730.1; --  
 DR PIR; A36911; A36911.  
 DR HSSP; P06201; ILGR.  
 DR InterPro; IPR001691; GLN synth.  
 DR InterPro; IPR004809; GLNA.  
 DR InterPro; IPR001637; GLNA adenyln.  
 DR Pfam; PF00120; Gln-synt; 1.  
 DR Pfam; PF03951; Gln-synt; N; 1.  
 DR ProDom; PD001057; Gln synt; C; 1.  
 DR TIGRFAMs; TIGR00653; GLNA; 1.  
 DR PROSITE; PS00180; GLNA.1; 1.  
 DR PROSITE; PS00181; GLNA\_ATP; 1.  
 KW Ligase.  
 FT BINDING 358 358 AMP (UNDER CONDITIONS OF ABUNDANT  
 GLUTAMINE). (BY SIMILARITY).  
 SQ SEQUENCE 439 AA; 50066 MW; C503B43B790BD25C CRC64;

Query Match 11.8%; Score 82; DB 1; Length 439;  
 Best Local Similarity 18.1%; Pred. No. 20;  
 Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

QY 25 GYQFDGWEISGEGKKGAGYVNLKDTFI-----KP-----VFK 59  
 DB 45 GISEFGSSVPGFQIGEDSLVFKADPDYVEVPWNVARVGYFYKDKPYGADPRGILK 104  
 QY 60 KIEEKEEBEN-----KPTFDVSKKKDN-----81  
 DB 105 RALELEKEGKYKAYIGPEPEFYLFKNGTWLEIPDVGGYFDILTLDKARDIRREIAEYM 164  
 QY 82 -----PQVNSQLNESHRKDLQREHSQKSD---STKDVATVLDKNNI 123  
 DB 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKTADNIVSFKYITKAVAEHGL 215

RESULT 22  
 ID IF2\_HELPFJ STANDARD; PRT; 949 AA.  
 AC Q9ZM46;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Translation initiation factor IF-2.  
 GN INF2 OR JHP0377.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=85963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99120557; PubMed=9923682;  
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
 RA Smith D.R., Noonan B., Gullid B.C., deJonge B.L., Carmel G.,  
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.P.,  
 RA Trust T.J.;  
 RT "Genomic sequence comparison of two unrelated isolates of the human  
 gastric pathogen Helicobacter pylori.";  
 RL Nature 397:176-180(1999).  
 CC -!- FUNCTION: One of the essential components for the initiation of  
 protein synthesis. Protects formylmethionyl-tRNA from spontaneous  
 hydrolysis and promotes its binding to the 30S ribosomal subunits.  
 CC Also involved in the hydrolysis of GTP during the formation of the  
 70S ribosomal complex (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AE001472; AAO05948.1; --  
 DR PIR; E71940; E71940.  
 DR HAMAP; MF\_00100; -; 1.  
 DR InterPro; IPR000795; EF GTPbind.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR000178; IF2\_D.  
 DR InterPro; IPR008847; IF2\_N.  
 DR InterPro; IPR005225; Small GTP.  
 DR Pfam; PF00009; GTP EFTU; 1.  
 DR Pfam; PF03144; GTP EFTU D2; 2.  
 DR Pfam; PF04760; IF2\_N; 1.  
 DR ProDom; PD186100; IF2; 1.  
 DR TIGRFAMs; TIGR00487; IF-2; 1.  
 DR TIGRFAMs; TIGR00231; small GTP; 1.  
 DR PROSITE; PS01176; IF2; 1.  
 KW Initiation factor; Protein biosynthesis; GTP-binding;  
 FT DOMAIN 451 599 G-DOMAIN.  
 FT NP\_BIND 457 464 GTP (BY SIMILARITY).



PT NP BIND 503 507 GTP (BY SIMILARITY).  
FT NP BIND 557 560 GTP (BY SIMILARITY).  
FT DOMAIN 321 326 POLY-GLU.  
SQ SEQUENCE 949 AA; 105961 MW; FA8969B0C64B3278 CRC64;  
  
Query Match 11.8%; Score 82; DB 1; Length 949;  
Best Local Similarity 28.7%; Pred. No. 45;  
Matches 29; Conservative 20; Mismatches 20; Indels 32; Gaps 5;  
  
Qy 50 KDTPIKVFVKIEBKKEENKPTFDVSKKD-----NPO-VNHSQLES 92  
Db 98 EETKAQPKPKTEKKKEAP-----IKKKEIIVTFENQTPDVNTKAVSHSQIEKA 154  
Qy 93 HRKEDLOREHSQKSDTKDVTATVLDKNKISSKSTNNPN 133  
Db 155 --KQKLOEIQKREA-----LNKLTQSNTNTNNAN 183  
  
RESULT 23  
LMD1 HUMAN  
ID LMD1 HUMAN STANDARD; PRT; 572 AA.  
AC P29536;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Leiomodlin 1 (Leiomodlin, muscle form) (64 kDa autoantigen D1) (64 kDa  
DE autoantigen ID) (64 kDa autoantigen ID3) (Thyroid-associated  
DE ophthalmopathy autoantigen) (Smooth muscle leiomodlin) (SM-Lmod).  
GN LMOD1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Thyroid;  
RX MEDLINE=91225220; PubMed=2026759;  
RA Dong Q., Ludgate M., Vassart G.;  
RT "Cloning and sequencing of a novel 64-kDa autoantigen recognized by  
RT patients with autoimmune thyroid disease.";  
RL J. Clin. Endocrinol. Metab. 72:1375-1381(1991).  
RN [2]  
RN CHARACTERIZATION.  
RX MEDLINE=99451105; PubMed=10520227;  
RA Conley C.A., Fowler V.M.;  
RT "Localization of the human 64kD autoantigen D1 to myofibrils in a  
RT subset of extraocular muscle fibers.";  
RL Curr. Eye Res. 19:313-322(1999).  
RN [3]  
RN CHARACTERIZATION.  
RX MEDLINE=21248187; PubMed=11350761;  
RA Conley C.A.;  
RT "Leiomodlin and tropomodulin in smooth muscle.";  
RL Am. J. Physiol. 280:C1645-C1656(2001).  
RN [4]  
RN CHARACTERIZATION.  
RX MEDLINE=21218919; PubMed=11318603;  
RA Conley C.A., Fritz-Six K.L., Almenar-Queralt A., Fowler V.M.;  
RT "Leiomodlins: larger members of the tropomodulin (Tmod) gene family.";  
RL Genomics 73:127-139(2001).  
CC -1- SUBCELLULAR LOCATION: CYTOSKELETON.  
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE (HEART, SKELETAL MUSCLE, COLON  
CC AND SMALL INTESTINE), A SUBSET OF STRIATED MUSCLE FIBERS, AND AT  
CC LOW LEVEL IN THYROID.  
CC -1- DISEASE: RECOGNIZED BY PATIENTS WITH AUTOIMMUNE THYROID DISEASE.  
CC -1- SIMILARITY: BELONGS TO THE TROPOMODULIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC EMBL; X54162; CAA38101.1; -.  
CC PIR; S18732; S18732.  
DR Genew; HGNC:6647; LMOD1.  
DR MIM; 602715; -.  
DR GO; GO:0005624; C:membrane fraction; TAS.  
DR GO; GO:0009405; P:pathogenesis; TAS.  
DR InterPro; IPR004934; Tropomodulin.  
DR InterPro; IPR003124; WH2.  
DR Pfam; PF03250; Tropomodulin; 1.  
DR Pfam; PF02205; WH2; 1.  
DR SMART; SM00246; WH2; 1.  
KW Antigen; Repeat; Cytoskeleton.  
FT DOMAIN 137 265 8 X APPROXIMATE TANDEM REPEATS.  
FT REPEAT 137 152 1.  
FT REPEAT 153 168 2.  
FT REPEAT 169 184 3.  
FT REPEAT 185 200 4.  
FT REPEAT 201 216 5.  
FT REPEAT 217 232 6.  
FT REPEAT 233 248 7.  
FT REPEAT 249 265 8.  
FT DOMAIN 480 499 5 X 4 AA APPROXIMATE TANDEM REPEATS.  
SQ SEQUENCE 572 AA; 63737 MW; D4B42F880523DE94 CRC64;  
  
Query Match 11.7%; Score 81.5; DB 1; Length 572;  
Best Local Similarity 22.8%; Pred. No. 29;  
Matches 38; Conservative 21; Mismatches 57; Indels 51; Gaps 6;  
  
Qy 19 KGELEKGYQFDGWEISGFEKG--KDAGYVINLSKDTFIKPVFKK-----IEEK 64  
Db 99 RGGKKSFSRDRDEAGSGSGEKPKEKIIRGIDKGRVRAAVDKKEAGKDGGRGERAVATK 158  
Qy 65 KEENK-----PTDVSKKKKNPQVNHLSHSHRKE----- 96  
Db 159 KEBEKKGDRNTGLSRDKKREEMKEVAKDEKVKERRNTDTRKEGEMKMRAGNT 218  
Qy 97 DLQREHSQK-----SDSTKDVAT-----VLDKNNISSKSTNNPNK 134  
Db 219 DMKKEDEKVRKGTNTDTRKDKDEKVKNEPLHEKEAKDSKTPTPEK 265  
  
RESULT 24  
DRS1 YEAST  
ID DRS1 YEAST STANDARD; PRT; 752 AA.  
AC P32892;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Probable ATP-dependent RNA helicase DRS1.  
GN DRS1 OR YLL008W OR L1345.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=93087480; PubMed=1454790;  
RA Ripmaster T.L., Vaughn G.P., Woolford J.L. Jr.;  
RT "A putative ATP-dependent RNA helicase involved in Saccharomycetes  
RT cerevisiae ribosome assembly.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135(1992).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=S288C / FY23;  
RX MEDLINE=96405918; PubMed=8810043;  
RA Moosga T., Zimmermann F.K.;  
RT "Sequence analysis of the CEN12 region of Saccharomycetes cerevisiae on  
RT a 43.7 kb fragment of chromosome XII including an open reading frame  
RT homologous to the human cystic fibrosis transmembrane conductance  
RT regulator protein CFTR.";  
RL Yeast 12:693-708(1996).

```

[3]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansgorge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heues-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wamburt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hohenseil J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:87-90(1997).
CC -!- FUNCTION: PROBABLE HELICASE INVOLVED IN RIBOSOME ASSEMBLY.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00683; AAA34666.1; -
CC EMBL; X91488; CAA62783.1; -
CC EMBL; 273113; CAA97452.1; -
CC PIR; S64750; S64750.
CC HSP; Q58083; 1HV8.
CC SGD; S0003931; DRS1.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR000629; DEAD_box.
CC InterPro; IPR001650; Helicase_C.
CC Pfam; PF00270; DEAD; 1.
CC Pfam; PF00271; helicase_C; 1.
CC SMART; SM00487; HELIC; 1.
CC SMART; SM00490; HELIC; 1.
CC PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
KW ATP-binding; RNA-binding; Helicase; Nuclear protein.
FT DOMAIN 170 190 POLY-GLU.
FT NP_BIND 275 282 ATP (POTENTIAL).
FT SITE 385 388 DEAD BOX.
FT CONFLICT 1 68 KKKGKNNKKYKSGEDNLDVHEDLD -> MTKKSLRLRL
FT RRSGRVRIIRKRLVRGITSMEFRTWT (IN REF. 1).
SQ SEQUENCE 752 AA; 84843 MW; 60747607A6E5E48 CRC64;
Query Match 11.7%; Score 81.5; DB 1; Length 752;
Best Local Similarity 21.0%; Pred. No. 38;
Matches 30; Conservative 26; Mismatches 48; Indels 39; Gaps 6;
QY 8 VSEDFILPVYKGELEKGYQFD-----GWEI--SGPEGKKDAGYVINLSKDTFI 54
Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
53 VSGNLDVHEDLDAGFKFDLADDTTSPQGNFNLAEAGESNKKDAEAFV--KKQVDL 110
QY 55 KPVFKK-----IESKKEENKPTDVSKKDN-----PQVNHSQLNE 91
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
111 DKTIIRKGLVKMAHIDSKQEETEKE-KVEKENDSDDEELAMDGFGMAPNNGDENOS 169
QY 92 SHRKDLQREHHSQKSDSTKDTV 114
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
170 EEEEEEEKEEEEEEEQEEMT 192
RESULT 25
MAPB RAT
ID MAPB RAT STANDARD; Q9ER21; Q9QW92; PRT; 2459 AA.
AC P15205; Q62958; Q9ER21; Q9QW92;

```

```

DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_Taxid=10116;
RN [1]
RX SEQUENCE OF 1-142 FROM N.A.
RP STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A. DOMAIN, AND INDUCTION.
RX STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
recombinant rat MAP 1B.";
RL Eur. J. Cell Biol. 57:66-74(1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RX TISSUE=Spinal cord;
RX MEDLINE=90059871; PubMed=2555150;
RA Rienitz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
RA Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
nervous system that is immunologically related to microtubule-
associated protein 5.";
RL EMBO J. 8:2879-2888(1989).
RN [4]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Nothias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
in rat CNS and PNS during development.";
RL J. Neurosci. Res. 49:319-332(1997).
CC -!- FUNCTION: The function of brain MAPs is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC -!- INDUCTION: By nerve growth factor.
CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (By similarity).
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: TO MAP1A.
CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC 2459) was originally described as neuraxin in Ref.3.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC -----
DR EMBL; U52950; AAB17068.1; -.
DR EMBL; X60370; CAC16162.1; -.
DR EMBL; X16623; CAA34620.1; ALT_SEQ.
DR PIR; A56577; A56577.
DR InterPro; IPR000102; MAP1B neuraxin.
DR Pfam; PF00414; MAP1B neuraxin; 10.
DR PROSITE; PS00230; MAP1B NEURAXIN; 8.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2459 MAP1 LIGHT CHAIN LCL.
FT REPEAT 1869 1885 MAP1B 1.
FT REPEAT 1886 1902 MAP1B 2.
FT REPEAT 1903 1919 MAP1B 3.
FT REPEAT 1920 1936 MAP1B 4.
FT REPEAT 1937 1953 MAP1B 5.
FT REPEAT 1954 1970 MAP1B 6.
FT REPEAT 1988 2004 MAP1B 7.
FT REPEAT 2005 2021 MAP1B 8.
FT REPEAT 2022 2038 MAP1B 9.
FT REPEAT 2039 2055 MAP1B 10.
FT DOMAIN 559 1035 GLU-RICH.
FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT LYS AND KEEI/V REPEATS).
FT DOMAIN 2224 2312 KEE AND KEEI/V REPEATS).
FT CONFLICT 127 127 M -> V (IN REF. 1).
FT CONFLICT 140 140 R -> S (IN REF. 1).
FT CONFLICT 2112 2112 R -> K (IN REF. 3).
FT CONFLICT 2169 2169 L -> I (IN REF. 3).
FT SEQUENCE 2459 AA; 269497 MW; 2E3F6872DEDB8BA2 CRC64;
Query Match 11.7%; Score 81.5; DB 1; Length 2459;
Best Local Similarity 27.8%; Pred. No. 1.3e+02;
Matches 25; Conservative 18; Mismatches 28; Indels 19; Gaps 4;
Qy 48 LSKDTFTKPVFK-KIEKKKEENKPTFDVSKKKNPQVNHSLNESHRKEDLQREHSQK 106
Db 631 VTDDKVVKKETKTPKEKKEE--KPKKEVAKKED-----KTPKKDKKPKK 674
Qy 107 SDSFKDVTATVL--DKNNISSKSTTNNPK 134
Db 675 EAKKEIKKEIKKEKKELKEVKKETPLK 704
RESULT 26
PAT1_YEAST
ID PAT1_YEAST STANDARD; PRT; 797 AA.
AC P25644;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Topoisomerase II-associated protein PAT1.
GN PAT1 OR YCR077C OR YCR77C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RA Ballesta J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,
RA Sanz E.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
[2]
RN REVISIONS.
RX MEDLINE=95373282; PubMed=7645349;
RA Rodriguez-Cousino N., Lill R., Neupert W., Court D.A.;
RT "Identification and initial characterization of the cytosolic protein
RT Ycr77p."
RT Yeast 11:581-585 (1995).
```

```
RN [3]
RP FUNCTION.
RX MEDLINE=971128274; PubMed=8972867;
RA Wang X., Watt P.M., Louis E.J., Borts R.H., Hickson I.D.;
RT "Pat1, a topoisomerase II-associated protein required for faithful
RT chromosome transmission in Saccharomyces cerevisiae.";
RL Nucleic Acids Res. 24:4791-4797 (1996).
CC -!- FUNCTION: NECESSARY FOR ACCURATE CHROMOSOME TRANSMISSION DURING
CC CELL DIVISION. INTERACTS WITH TOPOISOMERASE TOP2.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: TO S.POMBE SPBC19G7.10C.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL; X59720; CAA42266.1; -.
DR PIR; S53590; S53590.
DR SGD; S0000673; PAT1.
DR GO; GO:0005843; C:cytosolic small ribosomal subunit (sensu Eu. . .; IDA.
DR GO; GO:0007059; P:chromosome segregation; IMP.
DR GO; GO:0006446; P:regulation of translational initiation; IGI.
FT DOMAIN 195 200 POLY-PRO.
FT SEQUENCE 797 AA; 88487 MW; 2A54F39AE3E75ECE CRC64;
Query Match 11.6%; Score 81; DB 1; Length 797;
Best Local Similarity 25.8%; Pred. No. 45;
Matches 40; Conservative 21; Mismatches 58; Indels 36; Gaps 7;
Qy 5 STVSE-----EDFTLPVVK-----GELEKG-----YQFDGWEISGPGKKDAGTVI 46
Db 293 SQIVTEPDYNEDFYQVYKIQRGITSENKGLIARAYLEHSHRGLRGYKRTD----- 347
Qy 47 NLSKDTFTKPVFKKIEKKKEENK-----PTFDVSKKKNPQVNHSLNESHRKE 96
Db 348 -IALQRMQSQVERKAVTAKERPCKLKDQQAAGNSQDNKQANTVLGKISLTLSKNRPR 406
Qy 97 DLQ--REHSQKSDSTKDVATVLDKNNISKSTT 129
Db 407 QLQIPROQSPSSDPDALKDVTDSLNTVDLSSGSSS 441
RESULT 27
YCG1_YEAST
ID YCG1_YEAST STANDARD; PRT; 853 AA.
AC P25588; P25589; P27513; P87003;
DT 01-MAY-1992 (Rel. 22, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 97.9 kDa protein in CHAI-KRRI intergenic region.
GN YCL061C OR YCL61C/YCL60C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RA Rasmussen S.W., von Wettstein D.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
[2]
RN REVISIONS.
RA Gromadka R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 384-596 FROM N.A.
RC STRAIN=ATCC 28383 / FL100;
RX MEDLINE=90384830; PubMed=2169608;
RA Kern L.;
RT "The URK1 gene of Saccharomyces cerevisiae encoding uridine kinase.";
```

RC STRAIN=972;  
RC MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rayndream M.A., Lyne M., Lyne R., Stewart A.,  
RA Scourto J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth I., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feitwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidaoglou J., Hodgson G.,  
RA Helroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels J.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Waldfjens J., Volckaert G., Aart R., Robben J., Grymonprez B.,  
RA Wiedtjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzým K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,

Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).

•





RT "An essential yeast gene with homology to the exonuclease-encoding  
RT XRN1/KEM1 gene also encodes a protein with exoribonuclease  
RT activity.";  
RL Mol. Cell. Biol. 13:341-350(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93268292; PubMed=8497260;  
RA Aldrich T.L., di Segni G., McConaughy B.L., Keen N.J., Whelen S.,  
RA Hall B.D.;  
RA "Structure of the yeast TAP1 protein: dependence of transcription  
RT activation on the DNA context of the target gene.";  
RL Mol. Cell. Biol. 13:3434-3444(1993).  
RN [4]  
RN SEQUENCE FROM N.A.  
RP Landt O., Hiesel R., Unseid M.;  
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Bohn C., Bolotin-Fukuhara M., Daignan-Fornier B., Dang D.V.,  
RA Valens M.;  
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RL "- FUNCTION: May function in the processing and/or trafficking of  
CC nuclear mRNA. May be involved in general transcription as well.  
CC Possesses 5'->3' exoribonuclease activity degrading poly(A) to  
CC mainly 5'-AMP.  
CC "- COFACTOR: Requires magnesium.  
CC "- SUBCELLULAR LOCATION: Nuclear.  
CC "- SIMILARITY: Belongs to the 5'-3' exonuclease family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; S61567; AAB26818.1; -;  
DR EMBL; M95626; AAA34960.1; -;  
DR EMBL; L06011; AAA16950.1; -;  
DR EMBL; Z11746; -, NOT ANNOTATED\_CDS.  
DR EMBL; Z74956; CA99240.1; -;  
DR PIR; S20126; S20126.  
DR SGD; S0005574; RAT1.  
DR GO; GO:0005634; C:nucleus; IDA.  
DR GO; GO:0004534; F:5'-3' exoribonuclease activity; IDA.  
DR GO; GO:0006365; P:35S primary transcript processing; IMP.  
DR GO; GO:0006396; P:RNA processing; IMP.  
DR InterPro; IPR004859; Put\_53exo.  
DR Pfam; PF03159; XRN\_N; 1.  
KW Nuclear protein; Hydroxylase; Nuclease; Exonuclease; Repeat.  
FT DOMAIN 525 528 POLY-GLU.  
FT DOMAIN 793 797 POLY-ASN.  
FT DOMAIN 955 999 CONTAINS 2 X SRD, 2 X NNNY, AND 2 X  
FT YSGN REPEATS.  
FT MUTAGEN 683 Y->H: IN ALLELE TAP1-1; ACTIVATES  
FT TRANSCRIPTION OF THE PROMOTER-DEFECTIVE  
FT YEAST SUP4 TRNA(TYR) ALLELE SUP4A53T61.  
SQ SEQUENCE 1006 AA; 115933 MW; 5DD5B0245F3E12A CRC64;  
  
Query Match 11.6%; Score 80.5; DB 1; Length 1006;  
Best Local Similarity 20.0%; Pred. No. 62;  
Matches 34; Conservative 37; Mismatches 44; Indels 55; Gaps 9;  
  
Qy 1 KEMSS-----TIVSEEDFILPVY--KGELEKGYQFDGWEISGFGKKGAGYVNLKSDT 52  
Db 413 KWSKGDQRHTVATEQ---LQWYDTQGNLAKG-----SWNLTTSD-----MYRLKKEL 458  
Qy 53 FI-----KPVFKTIEKKBEENKPTFDVSKKKNPQVNHQ-----LNESH 93  
Db 459 MLANEGNEEALAKVQKQSDKNELMKDISKEIIDAVSKANKTNFLNLAEVNMQKIINKKH 518  
Qy 94 RKDLQREH-----SQKSUSTKDVATVLDK-----NNISSKSTT 129

Db 519 RLEKDNSEEBEIAKDSKVKTEKAESECDLDAETKDEIVADVNDRENSSETT 568  
  
RESULT 33  
LIP\_STAEP STANDARD; PRT; 688 AA.  
ID -LIP\_STAEP Q02510;  
AC Q02510;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Lipase precursor (EC 3.1.1.3) (Glycerol ester hydrolase).  
GN GEHC OR SE0281.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 303-315.  
RC STRAIN=9;  
RX MEDLINE=93171870; PubMed=8436947;  
RA Farrell A.M., Foster T.J., Holland K.T.;  
RT "Molecular analysis and expression of the lipase of Staphylococcus  
RT epidermidis.";  
RL J. Gen. Microbiol. 139:267-277(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
RA Chen Z., Wen Y.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
CC "- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
CC fatty acid anion.  
CC "- SUBCELLULAR LOCATION: Secreted.  
CC "- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS  
CC NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE  
CC CONVERSION).  
CC "- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M95577; AAA19729.1; -;  
DR EMBL; AE016744; AAO03878.1; -;  
DR PIR; A47705; A47705.  
DR InterPro; IPR005877; Gpos\_YSIK.  
DR InterPro; IPR000734; Lipase.  
DR InterPro; IPR000379; Ser\_estrs\_site.  
DR Pfam; PF04650; YSIK\_signal; 1.  
DR TIGRFAMs; TIGR01168; YSIK\_signal; 1.  
DR PROSITE; PS00120; LIPASE\_SER; 1.  
KW Hydroxylase; Lipid degradation; Zymogen; Signal; Complete proteome.  
FT SIGNAL 1 35 POTENTIAL.  
FT PROPEP 36 302 REMOVED IN THE MATURE FORM.  
FT CHAIN 303 688 LIPASE.  
FT ACT\_SITE 418 418 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 648 648 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CONFLICT 96 96 W->L (IN REF. 1).  
FT CONFLICT 120 120 E->G (IN REF. 1).  
SQ SEQUENCE 688 AA; 77343 MW; 6C95DB3A78AF86F6 CRC64;  
  
Query Match 11.5%; Score 80; DB 1; Length 688;  
Best Local Similarity 26.6%; Pred. No. 46;  
Matches 34; Conservative 24; Mismatches 50; Indels 20; Gaps 6;  
  
Qy 11 EDFILPVYKGELEKGYQFDGWEISGFGKKGAGYVNLKSDTFFKPVFKKIE-EKKEE 68  
Db 94 EDWI-----QQKQSDNDKNSVVEQNKENEAFVQNHSEE---KQOQEQVELEKHA 144

QY 69 NKPTEDVSKDNPOV--NHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSK 126  
 Db 145 NNQTLHSAAGNEDVKTKPSQLDNTAAKQDSQKSNLSKQDTQSSKTDLL-----R 197  
 QY 127 STTNPNK 134  
 Db 198 ATAQNSK 205

RESULT 34  
 ID IF2P HUMAN STANDARD; PRT; 1220 AA.  
 AC O60841; O95805; Q9UP81; Q9UMN7;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Translation initiation factor IF-2.  
 GN IF2 OR KIAA0741.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99087487; PubMed=9872452;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RA "Prediction of the coding sequences of unidentified human genes. XI.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 5:277-286 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cervical Carcinoma;  
 RX MEDLINE=99362399; PubMed=10432305;  
 RA Wilson S.A., Siero-Vazquez C., Edwards N.J., Lourin O., Byles E.D.,  
 RA Kotsopoulou E., Adamson C.S., Kingsman S.M., Kingsman A.J.,  
 RA Martin-Rendon E.;  
 RA "Cloning and characterization of hIF2, a human homologue of bacterial  
 RT translation initiation factor 2 and its interaction with HIV-1  
 RT matrix.";  
 RL Biochem. J. 342:97-103 (1999).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=Testis;  
 RX MEDLINE=99218282; PubMed=10200264;  
 RA Lee J.H., Choi S.K., Roll-Mecak A., Burley S.K., Dever T.E.;  
 RA "Universal conservation in translation initiation revealed by human  
 RT and archaeal homologs of bacterial translation initiation factor  
 RT IF2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:4342-4347 (1999).  
 RN [4]  
 RP SEQUENCE OF 89-1220 FROM N.A.  
 RC TISSUE=Testis;  
 RA Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;  
 RA Submitted (DEC-1999) to the EMBL/GenBank/DBSJ databases.  
 RN [5]  
 RP SEQUENCE OF 833-1220 FROM N.A.  
 RX MEDLINE=21064499; PubMed=11124703;  
 RA Stanchi F., Bertocco E., Toppi S., Dioguardi R., Simionati B.,  
 RA Cannata N., Zibello R., Lanfranchi G., Valle G.;  
 RA "Characterization of 16 novel human genes showing high similarity to  
 RT yeast sequences.";  
 RL Yeast 18:69-80 (2001).  
 CC -!- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING  
 CC THE BINDING OF THE FORMYL METHIONINE-TRNA TO RIBOSOMES. SEEMS TO  
 CC FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AB018284; BAA34461.1; -  
 DR EMBL; AJ006776; CAB44357.1; -  
 DR EMBL; AF078035; AAD16006.1; -  
 DR EMBL; AL133563; CAB63717.1; -  
 DR EMBL; AJ006412; CAA07018.1; -  
 DR PIR; T43483; T43483.  
 DR MIN; 606086; -  
 DR GO; GO:0003743; P:translation initiation factor activity; NAS.  
 DR GO; GO:0006446; P:regulation of translational initiation; NAS.  
 DR InterPro; IPR000795; EF\_GTPbind.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR000178; IF2.  
 DR InterPro; IPR005225; Small GTP.  
 DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
 DR PRINTS; PR00315; ELONGATNFT.  
 DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
 DR PROSITE; PS01176; IF2; FALSE\_NEG.  
 KW Initiation factor; Protein biosynthesis; GTP-binding.  
 FT NP\_BIND 638 645 GTP (BY SIMILARITY).  
 FT DOMAIN 39 50 POLY-LYS.  
 FT DOMAIN 94 99 POLY-LYS.  
 FT DOMAIN 138 142 POLY-ASP.  
 FT DOMAIN 313 322 POLY-LYS.  
 FT DOMAIN 353 356 POLY-GLU.  
 FT DOMAIN 361 364 POLY-GLU.  
 FT DOMAIN 491 496 POLY-GLU.  
 FT DOMAIN 529 567 ASP/GLU-RICH (ACIDIC).  
 FT MUTAGEN 640 640 V->G: LOSS OF ACTIVITY IN VIVO. RETAINS  
 FT MUTAGEN 706 706 FULL ACTIVITY IN VITRO.  
 FT MUTAGEN 706 706 H->E: LOSS OF ACTIVITY; BOTH IN VIVO AND  
 FT MUTAGEN 759 759 H->O: LOSS OF ACTIVITY IN VIVO. PARTIAL  
 FT MUTAGEN 759 759 D->N: LOSS OF ACTIVITY; BOTH IN VIVO AND  
 FT CONFLICT 64 64 IN VITRO.  
 FT CONFLICT 92 92 E -> G (IN REF. 1).  
 FT CONFLICT 180 180 T -> I (IN REF. 2).  
 FT CONFLICT 256 256 I -> M (IN REF. 3).  
 FT CONFLICT 522 522 K -> R (IN REF. 3).  
 FT CONFLICT 549 549 T -> K (IN REF. 3).  
 FT CONFLICT 549 549 E -> V (IN REF. 1).  
 FT CONFLICT 669 669 E -> W (IN REF. 3).  
 FT CONFLICT 894 894 E -> K (IN REF. 2).  
 SQ SEQUENCE 1220 AA; 138799 MW; 010CC15DE475E5B8 CRC64;  
 Query Match 11.5%; Score 80; DB 1; Length 1220;  
 Best Local Similarity 24.8%; Pred. No. 83;  
 Matches 35; Conservative 19; Mismatches 63; Indels 24; Gaps 4;  
 QY 10 BEDFILPVYKELEKGVQFDGWEISGEGKKDAGVINLSKDTFKPKVFKLEKKKESEN 69  
 Db 189 ESDEFLOSRRGQKKQKNKPGNIESGNEDDASFK-----IKTVAQKKAERER 240  
 QY 70 KPTFD---VSKKDNPOVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSK 113  
 Db 241 KKRDEKAKLRLKEKELETKGKQDSQKSEQRKFEETVSKVTVDGTGVPASEKAE 300  
 QY 114 TATVLDKNNISSKSTTNPNK 134  
 Db 301 TPTAAEDDNEGDKKKKKKKK 321

RESULT 35  
 ID CENC HUMAN STANDARD; PRT; 943 AA.  
 AC Q03188; Q9P0M5;  
 DT 01-OCT-1993 (Rel. 27, Created)



01-OCT-1993 (Rel. 27, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
Centromere protein C (CENP-C) (Centromere autoantigen C).  
CENPC1 OR CENPC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=92323541; PubMed=1339310;  
RA Satoh H., Tomkowiak J., Cooke C.A., Rathe H. III, Maurer M.,  
Rothfield N.F., Earnshaw W.C.;  
RT "CENP-C, an autoantigen in scleroderma, is a component of the human  
inner kinetochore plate.";  
RL Cell 70:115-125(1992).  
RN [2]  
RP SEQUENCE OF 1-21 FROM N.A.  
RC Poppe M., Botz J., Hamm B., Dobat K., Eickelbaum W., Pawletz N.,  
Arand M., Knehr M.;  
RA "Promoter characterization of centromere protein C reveals its  
participation in cell cycle regulation in late G1-phase and expression  
control by E2F-1, pRb, p107 and Sp-1.";  
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RL -!- FUNCTION: Component of the inner kinetochore plate. Required for  
normal kinetochore assembly.  
CC -!- SUBUNIT: Binds to DAXX.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M95724; AAA51974.1; -;  
DR EMBL; AF151723; AAF73191.1; -;  
DR PIR; A42681; A42681.  
DR Genbank; HGNC:1854; CENPC1.  
DR MIM; 117141; -;  
DR GO; GO:0005699; C:kinetochore; TAS.  
DR GO; GO:0003677; F:DNA binding activity; TAS.  
DR InterPro; IPR007113; Cupin\_sup.  
DR Chromosomal protein; Nuclear protein; DNA-binding; Phosphorylation;  
KW Centromere.  
KW DOMAIN 259 273 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 484 499 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 558 574 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 780 798 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD RES 75 75 PHOSPHORYLATION (POTENTIAL).  
FT MOD RES 732 732 PHOSPHORYLATION (POTENTIAL).  
SQ SEQUENCE 943 AA; 106925 MW; 6D80810A3E476376 CRC64;  
Query Match 11.4%; Score 79.5; DB 1; Length 943;  
Best Local Similarity 23.1%; Pred. No. 69;  
Matches 36; Conservative 16; Mismatches 55; Indels 49; Gaps 5;  
Qy 16 PVYGELEKGYQFDGWEISGFEKKDAGYVNLSDTKIPKVFVKIEKK-----ESEN 69  
Db 375 PSDKTLVTSYALIDETVNNYRSTKYEMYSKNAEKSFKRTIKQKQKRWKPAEAEQ- 433  
Qy 70 KPTFDVSKKDD-NPQVNHSQLNSHRKEDLQREEH-----SOKSGSTK 111  
Db 434 ---LDVQSGKDNIEHTSHITQDEFQNSDRNMEHEEENGVNCDVSKKPPVGVSKKSRK 490  
Qy 112 D-----VTATVLDKNNISK 126  
Db 491 DKESKKKRFSESKNKLVPPEVTSVTKSRISR 526

RESULT 36  
TRI DROME  
ID TRI DROME STANDARD; PRT; 268 AA.  
AC P36188; Q9VMV1; Q9VMV3; Q9VMV4;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Troponin I (TNI) (Wings apart-A protein) (Heldup protein).  
GN WUPA OR HDP OR TNI OR CG1718.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 9), AND FUNCTION.  
RC STRAIN=Canton-S; TISSUE=Embryo, and Larva;  
RX MEDLINE=91115093; PubMed=1892228;  
RA Barbas J.A., Gaiceran J., Krah-Jentgens I., de la Pompa J.L.,  
Canal I., Pongs O., Ferrus A.;  
RT "Troponin I is encoded in the haplolethal region of the Shaker gene  
complex of Drosophila.";  
RL Genes Dev. 5:132-140(1991).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 6).  
RC STRAIN=Oregon-R; TISSUE=Pupae;  
RX MEDLINE=91340840; PubMed=1908472;  
RA Beall C.J., Fyrberg E.;  
RT "Muscle abnormalities in Drosophila melanogaster heldup mutants are  
caused by missing or aberrant troponin-I isoforms.";  
RL J. Cell Biol. 114:941-951(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,  
Foaler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).

RN [4] REVISIONS, AND ALTERNATIVE SPLICING.  
RP STRAIN=Berkley;  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORM 8).  
RX STRAIN=Berkley; TISSUE=Head;  
RX MEDLINE=22426066; PubMed=12537569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,  
RA Rubin G.M., Celniker S.E.;  
RT "A Drosophila full-length cDNA resource.";  
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
RN [6]  
RP ALTERNATIVE SPLICING, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
RX MEDLINE=93180788; PubMed=7680094;  
RA Barbas J.A., Galceran J., Torroja L., Prado A., Ferrus A.;  
RT "Abnormal muscle development in the heldup3 mutant of Drosophila  
RT melanogaster is caused by a splicing defect affecting selected  
RT troponin I isoforms.";  
RL Mol. Cell. Biol. 13:1433-1439(1993).  
CC -!- FUNCTION: Troponin I is the ATPase inhibitory subunit of Troponin  
CC in the thin filament regulatory complex. Involved in the  
CC development and maintenance of muscle and nervous system. May also  
CC be involved in the cytoskeletal apparatus.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=10;  
CC Comment=Exon 3 is either present or absent, exon 6 has 4  
CC mutually exclusive forms (6a1, 6a2, 6b1 and 6b2) and C-terminal  
CC exons 9 and 10 are mutually exclusive;  
CC Name=10; Synonyms=G;  
CC IsoId=P36188-1; Sequence=Displayed;  
CC Name=1;  
CC IsoId=P36188-2; Sequence=VSP\_006626, VSP\_006627, VSP\_006630;  
CC Name=2; Synonyms=A;  
CC IsoId=P36188-3; Sequence=VSP\_006626, VSP\_006627;  
CC Name=3;  
CC IsoId=P36188-4; Sequence=VSP\_006626, VSP\_006628, VSP\_006630;  
CC Name=4; Synonyms=B;  
CC IsoId=P36188-5; Sequence=VSP\_006626, VSP\_006628;  
CC Name=5;  
CC IsoId=P36188-6; Sequence=VSP\_006626, VSP\_006630;  
CC Name=6; Synonyms=C, F;  
CC IsoId=P36188-7; Sequence=VSP\_006626;  
CC Name=7;  
CC IsoId=P36188-8; Sequence=VSP\_006626, VSP\_006629, VSP\_006630;  
CC Name=8; Synonyms=D, E;  
CC IsoId=P36188-9; Sequence=VSP\_006626, VSP\_006629;  
CC Name=9;  
CC IsoId=P36188-10; Sequence=VSP\_006630;  
CC -!- TISSUE SPECIFICITY: All isoforms are expressed in somatic muscle.  
CC Isoforms containing exon 6a1 (isoforms 1 and 2) are expressed in  
CC all muscles but highest expression is in abdominal muscle and  
CC splanchnic muscle of the gut. Isoforms containing exon 6b1  
CC (isoforms 5, 6, 9 and 10) are highly expressed in the tergal  
CC depressor of trochanter (TDT) muscle.  
CC -!- DEVELOPMENTAL STAGE: Isoforms containing exon 3 (isoform 9 and  
CC isoform 10) are expressed in adults. Isoforms containing exon 6a1  
CC (isoforms 1 and 2) are expressed at all developmental stages.  
CC Isoforms containing exon 6a2 (isoforms 3 and 4) are weakly  
CC expressed in embryos and larvae and very weakly in adults.  
CC Isoforms containing exon 6b1 (isoforms 5, 6, 9 and 10) are weakly  
CC expressed in larva and increase during metamorphosis. Isoforms

CC containing exon 6b2 (isoforms 7 and 8) are weakly expressed in  
CC embryos and larvae and at a higher level in adults.  
CC -!- MISCELLANEOUS: TROPONIN I BINDS TO ACTIN-TROPOMYOSIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X58188; CAA41171.1; -  
CC EMBL; X59376; CAA42020.1; -  
CC EMBL; AE003507; AAF48802.2; -  
CC EMBL; AE003507; AAF48803.2; -  
CC EMBL; AE003507; AAF48804.2; -  
CC EMBL; AE003507; AAF48805.2; -  
CC EMBL; AE003507; AAN09458.1; -  
CC EMBL; AY122145; AAM52657.1; -  
CC PIR; A40547; A40547.  
CC PIR; B38594; B38594.  
CC FlyBase: FBgn004028; wupa.  
CC GO; GO:0005861; C:troponin complex; NAS.  
CC GO; GO:0003779; F:actin binding activity; NAS.  
CC GO; GO:0007517; P:muscle development; IMP.  
CC GO; GO:0007399; P:neurogenesis; IMP.  
CC InterPro: IPR001978; Troponin.  
CC Pfam: PF00992; Troponin; I.  
CC Muscle protein; Actin-binding; Acetylation; Methylation;  
KW Alternative splicing.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
FT MOD\_RES 200 200 METHYLATION (TRI-) (BY SIMILARITY).  
FT MOD\_RES 204 204 METHYLATION (TRI-) (BY SIMILARITY).  
FT DOMAIN 161 170 TROPONIN T-INTERACTION (BY SIMILARITY).  
FT DOMAIN 188 201 ACTIN-BINDING (BY SIMILARITY).  
FT VARSPLIC 5 65 Missing (in isoform 1, isoform 2, isoform  
FT 3, isoform 4, isoform 5, isoform 6,  
FT 7 and isoform 8).  
FT VARSPLIC 150 182 /FTId=VSP\_006626.  
FT VARSPLIC 150 182 GELQICEYRYMYICGQKWDLEYEVKKDW -> DTLK  
FT SLIKQHYDRINKLDQKDYLVVVKRDV (in isoform  
FT 1 and isoform 2).  
FT VARSPLIC 150 182 /FTId=VSP\_006627.  
FT VARSPLIC 150 182 GELQICEYRYMYICGQKWDLEYEVKKDW -> DTIQ  
FT SVCKDVHSKILKLESEKYDFEYDVARKDY (in isoform  
FT 3 and isoform 4).  
FT VARSPLIC 150 182 /FTId=VSP\_006628.  
FT VARSPLIC 150 182 GELQICEYRYMYICGQKWDLEYEVKKDW -> AELQ  
FT TICKQYWRVYSLGDKPDLHVQKVKAAQ (in isoform  
FT 7 and isoform 8).  
FT VARSPLIC 248 268 /FTId=VSP\_006629.  
FT VARSPLIC 248 268 PDMSKGKPGDAKVEEVEAEA -> IKDAAVLNKKAKK (in  
FT isoform 1, isoform 3, isoform 5, isoform  
FT 7 and isoform 9).  
Query Match 11.4%; Score 79; DB 1; Length 268;  
Best Local Similarity 26.9%; Pred. No. 20;  
Matches 32; Conservative 19; Mismatches 56; Indels 12; Gaps 4;  
QY 10 BEDFTLPVYKGELEKGYQFDGWEISGFGKKDAGVIVNLKDTFKPVFKLEEKEEN 69  
Db 161 ERMYICGQKWDLEYEVKKDWEIN-----DLNAQVNDLRGKFKVPAKKV---SKYEN 211  
QY 70 KPTFDVSKKDNQVNHSHQSNESHRKE-DLQREHSHQSKSDSTKDVATVLDKNNISKS 127  
Db 212 K--FAKLOKAAEFNFRNLQVKKVKEFTLEEEKEKAPDWSKGPDAKVEEVEAEA 268  
RESULT 37  
REGA\_DICDI  
ID REGA\_DICDI STANDARD; PRT; 793 AA.

```
AC Q23917;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3',5'-cyclic-nucleotide phosphodiesterase regA (EC 3.1.4.17) (PDEase
DE regA).
OS REGA.
OS Dictyostelium discoideum (slime mold).
OS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Shaulesky G., Escalante R., Loomis W.F.;
RX "Developmental signal transduction pathways uncovered by genetic
RT suppressors.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RA Thomason P.A., Traynor D., Cavet G., Chang W.T., Harwood A.J.,
RA Kay R.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MORPHOLOGICAL SUPPRESSOR OF TAGS.
CC -!- CATALYTIC ACTIVITY: Nucleoside 3',5'-cyclic phosphate + H(2)O =
CC nucleoside 5'-phosphate.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT LOW LEVELS IN VEGETATIVE CELLS
CC AND AT HIGH LEVELS IN PRESPORE AND PRETALK CELLS DURING
CC DEVELOPMENT.
CC -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U60170; AAB03508.1; -.
CC EMBL; AJ005398; CAA06513.1; -.
CC Dictyob; DD02055; regA.
CC InterPro; IPR003607; Met_phosphohydro.
CC InterPro; IPR002073; PDEase.
CC Pfam; PF00233; PDEase; 1.
CC ProDom; PD000039; Response_reg.
CC PRINTS; PR00387; PDIESTERASE1.
CC ProDom; PD000039; Response_reg; 1.
CC SMART; SM00471; Hdc; 1.
CC SMART; SM00448; REC; 1.
CC PROSITE; PS01126; PDEASE_I; 1.
CC PROSITE; PS01110; RESPONSE_REGULATORY; 1.
KW Hydrolase; CAMP; CGMP.
FT DOMAIN 161 280 RESPONSE REGULATORY.
FT DOMAIN 18 28 POLY-SER.
FT DOMAIN 52 69 POLY-ASN.
FT DOMAIN 87 96 POLY-THR.
FT DOMAIN 100 121 POLY-ASN.
FT DOMAIN 166 169 POLY-ASP.
FT DOMAIN 770 776 POLY-SER.
SQ SEQUENCE 793 AA; 91175 MW; 68065A96620F8C27 CRC64;

Query Match 11.4%; Score 79; DB 1; Length 793;
Best Local Similarity 24.4%; Pred. No. 63;
Matches 21; Conservative 19; Mismatches 34; Indels 12; Gaps 2;

Qy 52 TFIKPVFKKIB-----EKKEENKPTFDVSKKNDPQVNHSQLNESHKREDLQREHSQKS 107
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 33 TSLKSMISGIEINLVHSGKNDN-----KNNNNNNNNNNNNKQKQDIVSLNNSSS 84
```

```
Qy 108 DSTKDVATVTLVDKNNISSKSTNNPN 133
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 NNITTTTTTTTTTSNHNNSNNNNNN 110

RESULT 38
Y372 AQUAE STANDARD; PRT; 531 AA.
AC O66695;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_372.
GN AQ_372.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE00686; AAC06652.1; -.
CC PIR; C70333; C70333.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 531 AA; 62830 MW; 6499DF803C2F7CF CRC64;

Query Match 11.3%; Score 78.5; DB 1; Length 531;
Best Local Similarity 25.3%; Pred. No. 45;
Matches 25; Conservative 17; Mismatches 34; Indels 23; Gaps 4;

Qy 3 MSSTIVSEEDFILPVYKGELEKGYQDPGWISGEGKKDAGYVNLKDTFIKPVFKKIE 62
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 MKEQILSKKLSIETLKNRLNPYTF---INGYD-----IISKFYRNSIF---- 107

Qy 63 EKKEENKPTFDVSKKNDPQVNHSQLNESHKREDLQRE 101
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 ----ANNPYKDI-KOENSKLLNSILNIDHRKEDMDYE 141

RESULT 39
CYLL HUMAN STANDARD; PRT; 598 AA.
AC P35663;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cyclicin I (Multiple-band polypeptide I) (Fragment).
GN CYLCI OR CYLI OR CYL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93359502; PubMed=8354692;
RA Hess H., Heid H., Franke W.W.;
```

```
RT "Molecular characterization of mammalian cylicin, a basic protein of
RT the sperm head cytoskeleton."
RL J. Cell Biol. 122:1043-1052(1993).
CC -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -!- TISSUE SPECIFICITY: Testis.
CC -!- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22780; CAA80457.1; -
DR PIR; B40713; B40713..
DR Genew; HGNC:2582; CYLC1.
DR MIM; 603121; -
DR GO; GO:0005856; C:cytoskeleton; NAS.
DR GO; GO:0005198; F:structural molecule activity; NAS.
DR GO; GO:0007283; P:spermatogenesis; NAS.
DR Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
KW NON_TER 1
FT DOMAIN 1 479 500 8 APPROXIMATE TANDEM REPEATS.
FT REPEAT 225 252 1.
FT REPEAT 253 289 2.
FT REPEAT 290 326 3.
FT REPEAT 327 364 4.
FT REPEAT 365 400 5.
FT REPEAT 401 438 6.
FT REPEAT 439 478 7.
FT REPEAT 479 500 8.
FT DOMAIN 548 598 PRO-RICH.
FT SEQUENCE 598 AA; 68034 MW; 062BA2E2D2AB61F7 CRC64;
SQ
Query Match 11.3%; Score 78.5; DB 1; Length 598;
Best Local Similarity 22.7%; Pred. No. 51;
Matches 32; Conservative 28; Mismatches 60; Indels 21; Gaps 8;
QY 5 STIVSE-----DFILPVYKGELEKGYQDQWEISGFEKGDAGYVINLS-----KDTFIKP 56
DB 182 SEICSENLSNVDFLMLVGQSD--DESINFDAW--LRNYSQNNKNYSLSKYTKYTKDT--KK 237
QY 57 VFKIEEKEENKPTFDVSKKNDPNVNHSQLNESHKEDLQREHS-----QKSDSTKDV 113
DB 238 NAKSSDAESEDK-----DAKDSKKVK-KNVKDDKKDKVKKDTESDGSKDER 291
QY 114 TATVLDKNNISSKSTNNPNK 134
DB 292 KDTKKDKKKLKKDKKDKTK 312
RESULT 40
DNL1 CABEL STANDARD; PRT; 847 AA.
AC Q27474;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA ligase (EC 6.5.1.1) (Polydeoxyribonucleotide synthase [ATP]).
GN LIG-1 OR C29A12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilkinson J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBSJ databases.
```

```
CC -!- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA
CC RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA.
CC -!- CATALYTIC ACTIVITY: ATP + {deoxyribonucleotide}(N) +
CC {deoxyribonucleotide}(M) = AMP + diphosphate +
CC {deoxyribonucleotide}(N+M).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z73970; CAA98242.1; -
DR PIR; T19544; T19544.
DR WormPep; C29A12.3; CE05328.
DR InterPro; IPR000977; DNA_ligase.
DR Pfam; PF01068; DNA_ligase_1.
DR Pfam; PF04675; DNA_ligase_A_C; 1.
DR Pfam; PF04675; DNA_ligase_A_N; 1.
DR TIGRFAMs; TIGR00574; dnll; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR PROSITE; PS00333; DNA_LIGASE_A2; 1.
DR PROSITE; PS0160; DNA_LIGASE_A3; 1.
KW DNA repair; DNA replication; DNA recombination; Cell division; Ligase;
KW ATP-binding.
FT BINDING 363 363 AMP (BY SIMILARITY).
SQ SEQUENCE 847 AA; 94501 MW; A32B7D279F91002F CRC64;
Query Match 11.3%; Score 78.5; DB 1; Length 847;
Best Local Similarity 25.2%; Pred. No. 74;
Matches 37; Conservative 24; Mismatches 59; Indels 27; Gaps 6;
QY 2 EMSSTIVSEEDFILPVYKGE-----LEKGYQDQWEISGFEKGDAGYVINL----- 48
DB 674 DKNSDDATSSQVLEMYKQAEAFANQIEKADAVD--EDDFEKEDEEELNMTNVSEG 731
QY 49 -SKDTFTKVPFKIEEKEENKPTFDVSKKK--DNFQVNHSQLNESHKEDLQREHSQ 105
DB 732 SSKENPVK-----EIKKETPK---SVSPKPKFEKPPVKSSPVNKPSPKSPKKEAK 782
QY 106 KSDSTKDV TATVLDKNNISSKSTNNP 132
DB 783 KKGVPASIFSSSTTKKNEKDVKVESFSP 809
Search completed: February 10, 2004, 10:49:50
Job time : 7.22948 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 23.725 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTIVSEBDFILPVYK.....ATVLKNNISKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

SPTREMBL 23:\*\*

- 1: sp\_archea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mhc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phase:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_rodent:\*\*
- 12: sp\_virus:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_unclassified:\*\*
- 15: sp\_rvirus:\*\*
- 16: sp\_bacteriap:\*\*
- 17: sp\_archaeap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	696	100.0	2119	2 Q9AHT5	Q9aht5 streptococc
2	696	100.0	2140	16 Q97RY6	Q97ry6 streptococc
3	693	99.6	2144	2 Q9S4W8	Q9s4w8 streptococc
4	693	99.6	2144	16 Q8DQ7	Q8dq7 streptococc
5	112	16.1	361	5 Q95P15	Q95p15 plasmodium
6	110	15.8	346	5 Q9U0G0	Q9u0g0 plasmodium
7	110	15.8	379	5 Q25705	Q25705 plasmodium
8	109	15.7	3008	5 Q81436	Q81436 plasmodium
9	108.5	15.6	600	5 Q77355	Q77355 plasmodium
10	108.5	15.6	1038	13 Q90784	Q90784 gallus gall
11	107	15.4	354	5 Q25995	Q25995 plasmodium
12	107	15.4	354	5 Q81J55	Q81j55 plasmodium
13	107	15.4	829	5 Q815F3	Q815f3 plasmodium
14	103.5	14.9	379	5 Q9U6C4	Q9u6c4 plasmodium
15	103.5	14.9	380	5 Q26019	Q26019 plasmodium
16	102.5	14.7	379	5 Q25706	Q25706 plasmodium

17	101.5	14.6	382	5 Q9V7J0	Q9v7j0 drosophila
18	101.5	14.6	556	5 Q9V7I9	Q9v7i9 drosophila
19	101.5	14.6	785	5 Q9GQ82	Q9gq82 drosophila
20	100	14.4	329	5 Q9NFV9	Q9nf9 plasmodium
21	100	14.4	1130	5 Q8IJ24	Q8ij24 plasmodium
22	99	14.2	211	5 P91488	P91488 caenorhabdi
23	99	14.2	1859	5 Q8IC27	Q8ic27 plasmodium
24	99	14.2	2563	5 Q813A0	Q813a0 plasmodium
25	98	14.1	157	5 Q9VQV0	Q9vqv0 drosophila
26	98	14.1	951	5 Q96229	Q96229 plasmodium
27	97.5	14.0	449	5 Q8IHW3	Q8ihw3 plasmodium
28	97.5	14.0	556	5 Q9S593	Q9s593 drosophila
29	95	13.6	622	16 Q98QA1	Q98qa1 mycoplasma
30	95	13.6	3026	5 Q8ILS9	Q8ils9 plasmodium
31	95	13.6	3193	5 Q8I590	Q8i590 plasmodium
32	94.5	13.6	312	16 Q9PPL5	Q9ppl5 campylobact
33	94.5	13.6	614	5 Q8TIX2	Q8tix2 dictyosteli
34	94	13.5	210	5 Q02124	Q02124 caenorhabdi
35	94	13.5	535	5 Q17595	Q17595 caenorhabdi
36	94	13.5	1434	5 Q8IJ13	Q8ij13 plasmodium
37	94	13.5	3455	5 Q8IKQ1	Q8ikq1 plasmodium
38	93	13.4	450	2 Q52006	Q52006 streptococc
39	93	13.4	455	5 Q8MN05	Q8mn05 dictyosteli
40	93	13.4	3127	5 Q8IDA0	Q8ida0 plasmodium
41	92.5	13.3	219	16 Q9XOM6	Q9xom6 thermotoga
42	92.5	13.3	325	5 Q44016	Q44016 dictyosteli
43	92.5	13.3	421	16 Q8EJQ7	Q8ejq7 shewanella
44	92.5	13.3	426	5 Q96300	Q96300 dictyosteli
45	92.5	13.3	1183	5 Q8IC11	Q8ic11 plasmodium

## ALIGNMENTS

## RESULT 1

Q9AHT5	ID	Q9AHT5	PRELIMINARY;	PRT;	2119 AA.
AC	Q9AHT5;				
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)			
DE	Serine protease (Fragment).				
GN	PTA.				
OS	Streptococcus pneumoniae.				
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_TaxID=1313;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=N4.				
RX	MEDLINE=1116976; PubMed=11179332;				
RA	Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,				
RA	Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,				
RA	Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,				
RA	Langermann S., Johnson S., Koenig S.;				
RT	"Use of a Whole Genome Approach To Identify Vaccine Molecules				
RT	Affording Protection Against Streptococcus pneumoniae Infection."				
RL	Infect. Immun. 69:1593-1598(2001).				
CC	-!- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY				
CC	AN AMIDE BOND (BY SIMILARITY).				
DR	EMBL; AF291699; AAK19159.1; -.				
DR	HSPSP; P00782; 2SPT.				
DR	MEROPS; S08.064; -.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR006192; LPXTG.				
DR	InterPro; IPR003137; PA.				
DR	InterPro; IPR000209; Peptidase_S8.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF00746; Gram_pos_anchor; 1.				
DR	Pfam; PF02225; PA; 1.				
DR	Pfam; PF00082; Peptidase_S8; 2.				
DR	PRINTS; PR00723; SUBTILISIN.				
DR	TIGRFAMs; TIGR01167; LPXTG_anchor; 1.				

DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Cell wall; Peptidoglycan-anchor; Protease.  
FT NON\_TER 1  
SQ SEQUENCE 2119 AA; 238226 MW; 51799B7F6B3960A6A CRC64;  
  
Query Match 100.0%; Score 696; DB 2; Length 2119;  
Best Local Similarity 100.0%; Pred. No. 1.4e-45;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVNLKDTFIKPVFKK 60  
DB 1952 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVNLKDTFIKPVFKK 2011  
  
QY 61 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120  
DB 2012 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2071  
  
QY 121 NNISKSSTNNPNK 134  
DB 2072 NNISKSSTNNPNK 2085  
  
RESULT 2  
Q97RY6 PRELIMINARY; PRT; 2140 AA.  
AC Q97RY6;  
DT 01-OCT-2001 (Tremblrel. 18, Created)  
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Serine protease, subtilase family.  
GN SP0641.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]\_  
RP SEQUENCE FROM N.A.  
RC STRAIN=TIGR4;  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";  
RL Science 293:498-506 (2001).  
RE EMBL; AE007373; AAK74791.1; -.  
DR MEROPS; S08.064; -.  
DR TIGR; SP0641; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Protease; Complete proteome.

SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;  
  
Query Match 100.0%; Score 696; DB 16; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 1.4e-45;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVNLKDTFIKPVFKK 60  
DB 1973 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVNLKDTFIKPVFKK 2032  
  
QY 61 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120  
DB 2033 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2092  
  
QY 121 NNISKSSTNNPNK 134  
DB 2093 NNISKSSTNNPNK 2106  
  
RESULT 3  
Q9S4M8 PRELIMINARY; PRT; 2144 AA.  
AC Q9S4M8;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Cell wall-associated serine proteinase PrtA precursor.  
GN PrtA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]\_  
RP SEQUENCE FROM N.A.  
RC STRAIN=3.B;  
RT Bethe G., ten Thoren E., Bongaerts R.J.M., Heinz H.-P., Zysk G.;  
RT "Cloning and sequencing of a novel surface protease of Streptococcus pneumoniae";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RE -!- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY AN AMIDE BOND (BY SIMILARITY).  
CC EMBL; AF127143; AAD48399.1; -.  
DR HSSP; P00782; 2SPT.  
DR MEROPS; S08.064; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGR; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Cell wall; Peptidoglycan-anchor; Signal.  
FT SIGNAL 19  
FT CHAIN 20 2144  
FT CELL\_WALL-ASSOCIATED SERINE PROTEINASE  
FT PrtA.  
SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
  
Query Match 99.6%; Score 693; DB 2; Length 2144;  
Best Local Similarity 99.3%; Pred. No. 2.4e-45;  
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVNLKDTFIKPVFKK 60  
DB 1977 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVNLKDTFIKPVFKK 2036

```
Qy 61 IEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 120
Db 2037 IEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 2096
Qy 121 NNISKSTNNPNK 134
Db 2097 NNISKSTNNPNK 2110

RESULT 4
QSDQP7
ID QSDQP7 PRELIMINARY; PRT; 2144 AA.
AC QSDQP7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN PrtA OR SP0361.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McLaren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008434; AAK9365.1; -.
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match 99.6%; Score 693; DB 16; Length 2144;
Best Local Similarity 99.3%; Pred. No. 2.4e-45;
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVYINLSKDTFIKPVFKK 60
Db 1977 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVYINLSKDTFIKPVFKK 2036

Qy 61 IEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 120
Db 2037 IEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 2096

Qy 121 NNISKSTNNPNK 134
Db 2097 NNISKSTNNPNK 2110

RESULT 5
Q95PI5
ID Q95PI5 PRELIMINARY; PRT; 361 AA.
AC Q95PI5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Merozoite surface protein 3 (fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RX SEQUENCE FROM N.A.
RP STRAIN=FVO;
RA Hiseeda H., Saul A., Long C.A., Miller L.H., Stowers A.W.;
RT "Merozoite Surface Protein 3 and Protection Against Malaria in Aotus
```

```
RT Monkeys.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY044180; AAK94780.1; -.
FT NON TER 1
FT NON TER 361
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match 16.1%; Score 112; DB 5; Length 361;
Best Local Similarity 23.9%; Pred. No. 0.41;
Matches 37; Conservative 28; Mismatches 48; Indels 42; Gaps 6;

Qy 18 YKGELEKGYQ-----FD---GWEISGF--EGKDGAG-----YVI 46
Db 165 YAGKVEKDYERAKNAYQKQAVLKAKEASSYDILGWFEFGGVPHEKKEENMLSHLYVS 224
Qy 47 NLSKDTFIKPVFKKIEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOK 106
Db 225 SKDKENISKENDVDLDE-KEEEAEETEELKEEKEETESEISEDEEEEEEEEEEE 283
Qy 107 SDSTKQVATVLDKN-----NISSKSTNN 131
Db 284 NDKKKEQEKESNNENNDQKKDMEANLISKNNNN 318

RESULT 6
Q9U0G0
ID Q9U0G0 PRELIMINARY; PRT; 346 AA.
AC Q9U0G0;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Merozoite surface protein 3 (fragment).
GN MSP3.
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=20416497; PubMed=10960178;
RA Okenu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in
RL Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252286; CAB65754.1; -.
FT NON TER 1
FT NON TER 346
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAPA010 CRC64;

Query Match 15.8%; Score 110; DB 5; Length 346;
Best Local Similarity 26.2%; Pred. No. 0.56;
Matches 34; Conservative 26; Mismatches 54; Indels 16; Gaps 5;

Qy 17 VYKGELEKGYQFD-GWEISGF--EGKDGAG-----YVINLSKDTFIKPVFKKIEKKEE 68
Db 184 VLKAKEASSYNYILGWFEFGGVPHEKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEE 242
Qy 69 NKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATV-----LDKN 121
Db 243 AEETGEQELEKKEETESEINEDEQEKEEEEEKEEENNDKKEQAKESQNDQKEDMEAQ 302

Qy 122 NISSKSTNN 131
Db 303 NLISKNNNN 312

RESULT 7
ID Q25705 PRELIMINARY; PRT; 379 AA.
AC Q25705;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Polymorphic antigen.
```

OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KI;  
RX MEDLINE=98156743; PubMed=9497029;  
RA McColl D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the  
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31(1997).  
DR EMBL; U08851; AAC47831.1; --  
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match 15.8%; Score 110; DB 5; Length 379;  
Best Local Similarity 23.48; Pred. No. 0.62;  
Matches 37; Conservative 30; Mismatches 41; Indels 50; Gaps 7;

QY 18 YKGELEKGYQ-----FD---GWEISGF--EGKKDAG-----YVI 46  
|::|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
DB 185 YAGVKVDYERAKNAYKANQAVLKAXEASSYDILGWFGGGVPEHKKEENMLSHLYVS 244  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
QY 47 NLSEKDTIFKPVFKIEBKKEEKNKPTFDVSKKNPQNHSQLN-----ESH 93  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
DB 245 SKDKENISKENDVDLDSE-KEEAEATEEELEEKNEETSETSEDEEEEEEEKKEENE 303  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
QY 94 RKEDLOREEHSDKSDTKDYVTATVLDDKNNISSKSTTN 131  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
DB 304 KKEQEKEQENNDNQKKMEA-----QNLISKNNNN 336  
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

RESULT 8  
Q8I436 PRELIMINARY; PRT: 3008 AA.  
ID Q8I436  
AC Q8I436;  
DT DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF0325W.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Davlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,  
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;  
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22255708; PubMed=12368867;  
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
RA Bucke C.O., Burrows C., Cherevach I., Chillingworth C.,  
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,  
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,  
RA Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitsch E.,  
RA Rajadream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
RA Sulston J.E., Craig A., Newbold C., Barrell B.G;  
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
RL Nature 419:527-531(2002).  
DR EMBL; AL929351; CAD51431.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 3008 AA; 356023 MW; 60BCBEEE15C599B4 CRC64;

Query Match 15.7%; Score 109; DB 5; Length 3008;



DE Claustarin.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCB1\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Cole G.J.;  
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-451 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=94157526; PubMed=7906711;  
RA Burg M.A., Cole G.J.;  
RT "Claustarin, an antiadhesive neural keratan sulfate proteoglycan, is structurally related to MABP.B.";  
RL J. Neurobiol. 25:1-22(1994).  
DR EMBL; X67778; CAA47988.1; -  
SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;  
  
Query Match 15.6%; Score 108.5; DB 13; Length 1038;  
Best Local Similarity 28.6%; Pred. No. 2.3;  
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;  
  
Qy 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVINLSKDTFTKPVFK- 59  
Db 595 KPETKIVAEKDV-----TTKEQKGKSETSEKQASEKQDVKPKVKTSKVKKEVKA 646  
  
Qy 60 KIEKKEBENKPTFDVSKKNDPNQVNHSQLNESHKEDLQRE-----EHSQKSDSTKDV 113  
Db 647 KPKEKKDKEPKPKESVSKKEKPLI---KKEKPKKEDIKKEVKKEKKEKKEKKEV 702  
  
RESULT 11  
Q25995 ID Q25995 PRELIMINARY; PRT; 354 AA.  
AC Q25995;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Antigen.  
OS Plasmodium falciparum (isolate NF54).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCB1\_TaxID=5843;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NF54;  
RX MEDLINE=95198774; PubMed=7891748;  
RA McColl D.J., Silva A., Poley M., Kun J.F., Favaloro J.M.,  
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;  
RT "Molecular variation in a novel polymorphic antigen associated with Plasmodium falciparum merozoites.";  
RL Mol. Biochem. Parasitol. 68:53-67(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NF54;  
RX MEDLINE=98156743; PubMed=9497029;  
RA McColl D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31(1997).  
DR EMBL; L28825; AAC09377.1; -  
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;  
  
Query Match 15.4%; Score 107; DB 5; Length 354;  
Best Local Similarity 23.8%; Pred. No. 0.98;  
Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;  
  
Qy 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 53  
Db 185 KEASS-----YDYIL-----GWFGGVPPEHKKEENMLSHLYVSSKDKENI 225  
  
RESULT 12  
Q81J55 ID Q81J55 PRELIMINARY; PRT; 354 AA.  
AC Q81J55;  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Merozoite surface protein 3.  
GN PF10\_0345.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCB1\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=2225705; PubMed=12368864;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
RA Perteira M., Allen J., Salengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrell B.;  
RT "Genome sequence of the human malaria parasite Plasmodium falciparum.";  
RL Nature 419:498-511(2002).  
DR EMBL; AE014834; AAN35542.1; -  
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;  
  
Query Match 15.4%; Score 107; DB 5; Length 354;  
Best Local Similarity 23.8%; Pred. No. 0.98;  
Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;  
  
Qy 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 53  
Db 185 KEASS-----YDYIL-----GWFGGVPPEHKKEENMLSHLYVSSKDKENI 225  
  
Qy 54 IKPVFKKIEKKEE-----ENKPTFDVSKKNDPNQVNHSQLNESHKEDLQ 100  
Db 226 SKENDVDLDEKEEAEETEEELKEEKEETEESEISEDEEEEEEKEEENDKKKEQEK 285  
  
Qy 101 EEHSQKSDSTKDVATATVLDKNNISSKSTTN 131  
Db 286 EQSNNNDQKDMA-----QNLISKNNNN 311  
  
RESULT 13  
Q815F3 ID Q815F3 PRELIMINARY; PRT; 829 AA.  
AC Q815F3;  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF1275C.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCB1\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;

Qy 54 IKPVFKKIEKKEE-----ENKPTFDVSKKNDPNQVNHSQLNESHKEDLQ 100  
Db 226 SKENDVDLDEKEEAEETEEELKEEKEETEESEISEDEEEEEEKEEENDKKKEQEK 285  
  
Qy 101 EEHSQKSDSTKDVATATVLDKNNISSKSTTN 131  
Db 286 EQSNNNDQKDMA-----QNLISKNNNN 311  
  
RESULT 12  
Q81J55 ID Q81J55 PRELIMINARY; PRT; 354 AA.  
AC Q81J55;  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Merozoite surface protein 3.  
GN PF10\_0345.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCB1\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=2225705; PubMed=12368864;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
RA Perteira M., Allen J., Salengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrell B.;  
RT "Genome sequence of the human malaria parasite Plasmodium falciparum.";  
RL Nature 419:498-511(2002).  
DR EMBL; AE014834; AAN35542.1; -  
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;  
  
Query Match 15.4%; Score 107; DB 5; Length 354;  
Best Local Similarity 23.8%; Pred. No. 0.98;  
Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;  
  
Qy 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 53  
Db 185 KEASS-----YDYIL-----GWFGGVPPEHKKEENMLSHLYVSSKDKENI 225  
  
Qy 54 IKPVFKKIEKKEE-----ENKPTFDVSKKNDPNQVNHSQLNESHKEDLQ 100  
Db 226 SKENDVDLDEKEEAEETEEELKEEKEETEESEISEDEEEEEEKEEENDKKKEQEK 285  
  
Qy 101 EEHSQKSDSTKDVATATVLDKNNISSKSTTN 131  
Db 286 EQSNNNDQKDMA-----QNLISKNNNN 311  
  
RESULT 13  
Q815F3 ID Q815F3 PRELIMINARY; PRT; 829 AA.  
AC Q815F3;  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF1275C.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCB1\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;







RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RN Science 287:2185-2195(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banson J., An H., Baldwin D., Banton J., Besson K.Y., Buesam D.A.,  
 RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragaa V., Park S., Patel S., Pfeiffer B., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF289493; AAG40806.1; -;  
 DR EMBL; AE003808; AAM70947.1; -;  
 DR FlyBase; FBGN0034075; AspH.  
 DR InterPro; IPR001440; TPR.  
 DR InterPro; IPR006025; Zn MTpeptidase.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 SQ SEQUENCE 785 AA; 89843 MW; 30A8DFCD6836F7F1 CRC64;

Query Match 14.6%; Score 101.5; DB 5; Length 785;

Best Local Similarity 24.5%; Pred. No. 6;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 11 EDRFLPVYKGLKGYQFDGW-----EISFEKKGADAGYVI-----NLSKDTFK 55  
 DB 78 EDLDTPLSESRFSK--VFDGWVDHREHGDHGVQFSGEALDDHDEHDDHDEDEDEE 135  
 QY 56 PVPKKIEKEEENKPT-----FDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDS 109  
 DB 136 PLTEELSEELSEEEPTDEPADEVEDEDEENNA--CENTAEDAEDEEEDND 193  
 QY 110 TKQVATVLDKNNISSKST 128  
 DB 194 EGTVEATVEATTEATTEAT 212  
 RESULT 20  
 Q9NFV9 PRELIMINARY; PRT; 329 AA.  
 ID Q9NFV9 AC Q9NFV9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Merozoite surface protein 3 (fragment).  
 GN MSP3.  
 OS Plasmodium falciparum (isolate 7G8).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=57266;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=7G8;  
 RX MEDLINE=20416497; PubMed=10960178;  
 RA Okeno D.M.N., Thomas A.W., Conway D.J.;  
 RT "Allelic lineages of the merozoite surface protein 3 gene in  
 RT Plasmodium reichenowi and Plasmodium falciparum.";  
 RL Mol. Biochem. Parasitol. 109:185-188 (2000).  
 DR EMBL; AJ252287; CAB85901.1; -;  
 FT NON TER 1  
 FT NON TER 329  
 SQ SEQUENCE 329 AA; 36916 MW; C5B045DB5E21A159 CRC64;  
 Query Match 14.4%; Score 100; DB 5; Length 329;  
 Best Local Similarity 26.0%; Pred. No. 3.2;  
 Matches 38; Conservative 24; Mismatches 44; Indels 40; Gaps 7;  
 QY 1 KEMSSSTIVSEEDFLPVYKGLKGYQFDGWIEISGF--EGKKDAG-----TVINLSKDTF 53  
 DB 175 KEASS-----YDYIL-----GWFGGVPPEHKKEENMLSHLYVSSKDKENI 215  
 QY 54 IKPVFKKIEKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREH-----SQ 105  
 DB 216 SKENDVDVLDK-KESEAEETEEELKEEETSEISEDEDEEKEEKEEKEQKEQSGNE 274  
 QY 106 KSDTKQVATVLDKNNISSKSTNN 131  
 DB 275 NNDQKDMEA-----QNLISKNNNN 295  
 RESULT 21  
 Q8IJZ4 PRELIMINARY; PRT; 1130 AA.  
 ID Q8IJZ4 AC Q8IJZ4;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Hypothetical protein.  
 GN PF10\_0046.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;

[illegible]







```
Db 194 EGTVEATVEATTEATTEAT 212

RESULT 29
Q98QA1 PRELIMINARY; PRT; 622 AA.
AC Q98QA1
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE LIPOPROTEIN.
GN MYPU_4650.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OC NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "the complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL; AL445564; CAC13638.1; -
DR MyPUList; MYPU 4650; -
KW Complete proteome.
SQ SEQUENCE 622 AA; 73762 MW; 42BD88930861960D CRC64;

Query Match 13.6%; Score 95; DB 16; Length 622;
Best Local Similarity 25.4%; Pred. No. 15;
Matches 34; Conservative 26; Mismatches 50; Indels 24; Gaps 5;

Qy 8 VSEDFILPVVKGLEKGYQPDGWEISGFEGKKDAGYVILNSKDTFKVPFKIEKKKE 67
Db 85 LSEDDIIIFSLINENAKFQIDEF-----VSKDEKFKIKFQEIFSQTE 128
Qy 68 ENKPTFDVSKKKD---NPOVNHSQLNES--HRKEDLQREHSQKSDSTKDVTATVLDKN 121
Db 129 Q-KITDNISSKEDSKNNPKDNENSNNSDQKNDLQKNSDKLNDNVQDEKANKNSN 187
Qy 122 -NISSKSTTNPNK 134
Db 188 SNDSKKNENTNK 201

RESULT 30
Q8ILS9 PRELIMINARY; PRT; 3026 AA.
AC Q8ILS9
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF14_0165.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteau M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014849; AAN36406.1; -
KW Hypothetical protein.
SQ SEQUENCE 3193 AA; 377923 MW; 46B6B21F921C5307 CRC64;

Query Match 13.6%; Score 95; DB 5; Length 3193;
Best Local Similarity 28.3%; Pred. No. 83;
Matches 28; Conservative 21; Mismatches 36; Indels 14; Gaps 3;

Qy 38 GKXDAGYVILNSKDT-----ETKPVFKIEKKKEE--NKPTFDVSKKKNPQV 84
Db 1705 GKKSANSINGIKTSTCNCIHNDMKFKQYINFLIAQKEKINKEKINEKEKNDG 1764
Qy 85 NKSQNESHRKEDLQREHSQKSDSTKDVTATVLDKNKI 123
Db 1765 NKXE-ESSQKEESQKEESQKEESKNEESKNEENNI 1802

RESULT 32
Q9PPL5 PRELIMINARY; PRT; 312 AA.
AC Q9PPL5
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative membrane protein.
GN CU0692C.
```

```
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014818; AAN36777.1; -
KW Hypothetical protein.
SQ SEQUENCE 3026 AA; 357633 MW; 9ECAED915C325CE CRC64;

Query Match 13.6%; Score 95; DB 5; Length 3026;
Best Local Similarity 30.2%; Pred. No. 79;
Matches 29; Conservative 15; Mismatches 38; Indels 14; Gaps 4;

Qy 52 TFIKPVFKIEKKKEENKPTFDVSKKKD---NPOVNHSQLNESHRKEDLQREHSQKS 107
Db 235 TWANANKKJKNQVEEYKKNVSVINRDKNTIQNSDINNTQ-NILHRNDEIEEKLND 293
Qy 108 -DSTKQVTVATVLDKNNISSTKT-----NNPNK 134
Db 294 IHNVTKITKEYSSNSPSSNSDITLSYESVNNKKNK 329

RESULT 31
Q8I590 PRELIMINARY; PRT; 3193 AA.
AC Q8I590
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF11600C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteau M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014849; AAN36406.1; -
KW Hypothetical protein.
SQ SEQUENCE 3193 AA; 377923 MW; 46B6B21F921C5307 CRC64;

Query Match 13.6%; Score 95; DB 5; Length 3193;
Best Local Similarity 28.3%; Pred. No. 83;
Matches 28; Conservative 21; Mismatches 36; Indels 14; Gaps 3;

Qy 38 GKXDAGYVILNSKDT-----ETKPVFKIEKKKEE--NKPTFDVSKKKNPQV 84
Db 1705 GKKSANSINGIKTSTCNCIHNDMKFKQYINFLIAQKEKINKEKINEKEKNDG 1764
Qy 85 NKSQNESHRKEDLQREHSQKSDSTKDVTATVLDKNKI 123
Db 1765 NKXE-ESSQKEESQKEESQKEESKNEESKNEENNI 1802

RESULT 32
Q9PPL5 PRELIMINARY; PRT; 312 AA.
AC Q9PPL5
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative membrane protein.
GN CU0692C.
```

```
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteriaceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168; PubMed=10688204;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668 (2000).
RW EMBL; AL139076; CAB72966.1; -.
KW Complete proteome.
SQ SEQUENCE 312 AA; 37221 MW; 0004FA7836A741E8 CRC64;

Query Match 13.6%; Score 94.5; DB 16; Length 312;
Best Local Similarity 25.7%; Pred. No. 8;
Matches 38; Conservative 30; Mismatches 55; Indels 25; Gaps 7;

QY 8 VSEDFILPVYK-----GELEKGYQFDGWEISGFEKGDAGVIVNL---SKDTFTKP 56
DB 38 ISSDILRRFRKKTPTNKFLEEDDEYKSKHTKSNIVLKED---LINVLEEKOSLAKK 94
QY 57 VFKTIEEKEENKPT---FDVSKKKDNP---QVNHSQLNBSHRKDLQREHSQKSDS 109
DB 95 IFSKMKRERKEENKKTKNFLPSRKANEIKNIQTQIQTKSNQATTTQKQKELTNS 154

QY 110 TKDV--TATVLDKNNISK--STNNPN 133
DB 155 IEKIQTETKIQKPLIIEKKLDVKNQPN 182

RESULT 33
Q8TIX2 PRELIMINARY; PRT; 614 AA.
AC Q8TIX2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 71.6 kDa protein.
OC Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBSJ databases.
RW EMBL; AC116032; AAL93046.1; -.
KW Hypothetical protein.
SQ SEQUENCE 614 AA; 71591 MW; 85520D7AF08530F5 CRC64;

Query Match 13.6%; Score 94.5; DB 5; Length 614;
Best Local Similarity 23.9%; Pred. No. 16;
Matches 21; Conservative 23; Mismatches 43; Indels 1; Gaps 1;

QY 47 NLSKDTFTKVPKKTIEEKEENKPTFDVSKKKDNPQVNHSQLNBSHRKDLQREHSOK 106
DB 509 NLNKNISIRKEFVN-EIKNEKEVEKENKNEKNENENKNDENKNDEKNEE 567

QY 107 SDSTKDVTTATVLDKNNISKSTNNPNK 134
DB 568 DDDVIEIQPILKKQKTSSTSSVPKKNQ 595
```

```
RESULT 34
O02124 PRELIMINARY; PRT; 210 AA.
ID O02124
AC O02124;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E03H12.5 protein.
GN E03H12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Nelson, J., Wohlmann, P., Sansone J.;
RT "The sequence of C. elegans cosmid E03H12.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF000299; AAC47980.1; -.
DR WormPep; E03H12.5; CE09144.
SQ SEQUENCE 210 AA; 23703 MW; 641B64BCF7AC98B8 CRC64;

Query Match 13.5%; Score 94; DB 5; Length 210;
Best Local Similarity 28.6%; Pred. No. 5.8;
Matches 28; Conservative 18; Mismatches 52; Indels 0; Gaps 0;

QY 37 EGKDGAGVIVNLKDTFTKVPKKTIEEKEENKPTFDVSKKKDNPQVNHSQLNBSHRKE 96
DB 69 EGKDGKDGKSEKKGDKKEEKKDEKDEKDGKDEKDEKDEKDEKDEKDEKDEKDEKDE 128
QY 97 DLQREHSQKSDSTKDVTTATVLDKNNISKSTNNPNK 134
DB 129 EKDKDKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEK 166

RESULT 35
Q17595 PRELIMINARY; PRT; 535 AA.
ID Q17595
AC Q17595;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 60.3 kDa protein.
GN C02H7.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
```

[illegible]

```
[2]
RN SEQUENCE FROM N.A.
RP SPECIES=E.faecalis; STRAIN=RE25; PLASMID=PRE25;
RX PubMed=11735367;
RT Schwarz F.V., Perreten V., Teuber M.;
RT "Sequence of the 50-kb conjugative multiresistance plasmid PRE25 from
RT Enterococcus faecalis RE25.";
RL Plasmid 46:170-187(2001).
DR EMBL; L39769; AAA99471.1; -
DR EMBL; X92945; CAC29184.1; -
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 450 AA; 52755 MW; F3CA5C827CFF30CB CRC64;

Query Match 13.4%; Score 93; DB 2; Length 450;
Best Local Similarity 26.0%; Pred. No. 15;
Matches 39; Conservative 28; Mismatches 57; Indels 26; Gaps 7;

QY 9 SEEDFLPVYK-GELEK-----GYQFDGWEISGEGKKGAGVYNLSKDTFIKPVF 58
DB 302 SDKKIQSLYKQGETEKIEALDLPQSDYLADPKKILAYDNDSTIEYLVOTSTDTTI--VE 359
QY 59 KKEIEKEEENKPT-----FDVSKKK---DNPQVNHSQLNESHKEDLQR--BEHSQK 106
DB 360 ALIDSKVKEKDPTVKNRQVSIQTKKLAIDPKRQISMLDILLIENNESELENWYKDSINL 419
QY 107 SOSTKDVATVLDKNN--ISSKSTNNPNK 134
DB 420 EDLKKQKQNDLLEENNTLLSEKIETNAEK 449

RESULT 39
QBN05 PRELIMINARY; PRT; 455 AA.
AC Q8MN05;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Fachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC117177; AAM43671.1; -
KW Hypothetical protein.
SQ SEQUENCE 455 AA; 51270 MW; DBB00463D1FEA2A4 CRC64;

Query Match 13.4%; Score 93; DB 5; Length 455;
Best Local Similarity 32.9%; Pred. No. 16;
Matches 24; Conservative 10; Mismatches 33; Indels 6; Gaps 1;

QY 62 EEKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKN 121
DB 201 EEDEEVKVLMDIDTDTTKSKQKQKE-----NEDDEEEEGDIKSTKDLNLLVNNN 254
QY 122 NISSKSTNNPNK 134
DB 255 NSSITSSLYPNK 267

RESULT 40
Q8IDA0 PRELIMINARY; PRT; 3127 AA.
ID Q8IDA0;
AC Q8IDA0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Guanylyl cyclase (EC 4.6.1.2).
GN GC-BETA.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA Berriman M., Pain A., Hall N., Arkin R., Chillingworth C., Doggett J.,
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrall B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844509; CAD52725.1; -
KW Lyase.
SQ SEQUENCE 3127 AA; 370597 MW; F0375C72B9007560 CRC64;

Query Match 13.4%; Score 93; DB 5; Length 3127;
Best Local Similarity 24.7%; Pred. No. 1.2e+02;
Matches 45; Conservative 29; Mismatches 58; Indels 50; Gaps 7;

QY 1 KEMGSTIVSEEDFILPVYKGELEKGYQFDGWEISG---FEGKKDAGYVYNLSKDTFIKPV 57
DB 2000 KKEIKTIVSDDMFTSPVNIKEYNYNEQERKKEIVGNLSYDKTKKIFPFKFTKEGRKK- 2058
QY 58 FKKIEEKEEENK-----PTF-----DVSKKKNQV 84
DB 2059 -KKIEKKEKKEKKNNNNNFLYNDYSSYSSPKYGDNNNNFVIKYIRERKDKQKFDHNF 2117
QY 85 NHSQL-----NESHK-----EDLQREHSQKSDSTKDVATVLDKNISSKSTNNP 132
DB 2118 NFSKFLHYNPMKKNKKNKKNKNNRNEYPNYNTSSSKDGVSYNFLSDLSFSSDNEYSSD 2177
QY 133 NK 134
DB 2178 NE 2179

Search completed: February 10, 2004, 10:57:04
Job time : 24.725 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:57:15 ; Search time 27.9011 Seconds  
(without alignments)  
1230.730 Million cell updates/sec

Title: us-10-067-385-8\_copy\_610\_773

Perfect score: 848

Sequence: 1 TTVKETILNDTGEVSELKP.....ATVLDKNNISSKSTNNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/PCT PUBCOMB.pap.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT NEW PUB.pap.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06 NEW PUB.pap.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06 PUBCOMB.pap.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07 NEW PUB.pap.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS PUBCOMB.pap.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08 NEW PUB.pap.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08 PUBCOMB.pap.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A PUBCOMB.pap.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B PUBCOMB.pap.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C PUBCOMB.pap.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09 NEW PUB.pap.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A PUBCOMB.pap.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B PUBCOMB.pap.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C PUBCOMB.pap.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10 NEW PUB.pap.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60 NEW PUB.pap.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60 PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	773	14	US-10-067-385-8
2	848	100.0	2119	12	US-09-769-744A-28
3	615	72.5	117	9	US-09-765-272-68
4	106	12.5	665	11	US-09-820-843A-107
5	96.5	11.4	2060	12	US-10-381-596A-2
6	93.5	11.0	654	12	US-10-172-502-10
7	92	10.8	1875	12	US-10-369-493-22285
8	90.5	10.7	635	12	US-10-032-585-7046
9	90.5	10.7	743	15	US-10-171-311-188
10	90.5	10.7	758	9	US-09-925-299-859
11	90.5	10.7	758	11	US-09-925-299-859
12	89.5	10.6	402	10	US-09-827-664-4
13	89.5	10.6	511	12	US-10-289-762-509
14	89.5	10.6	516	12	US-10-032-585-7407
15	89.5	10.6	529	10	US-09-827-664-2

16	88	10.4	225	12	US-10-032-585-7829	Sequence 7829, Ap
17	88	10.4	1298	12	US-10-144-194A-106	Sequence 106, Appl
18	88	10.4	3051	12	US-10-144-194A-62	Sequence 62, Appl
19	87.5	10.3	655	12	US-10-369-493-2016	Sequence 2016, Ap
20	87.5	10.3	1349	9	US-09-815-242-5898	Sequence 5898, Ap
21	87.5	10.3	1349	9	US-09-815-242-13137	Sequence 13137, A
22	87.5	10.3	1363	12	US-10-032-585-7800	Sequence 7800, Ap
23	87.5	10.3	1702	11	US-09-839-996-5	Sequence 5, Appli
24	87.5	10.3	1702	15	US-10-080-505-5	Sequence 16456, A
25	87	10.3	323	12	US-10-369-493-16456	Sequence 21889, A
26	87	10.3	778	12	US-10-369-493-21889	Sequence 21889, A
27	87	10.3	6642	12	US-10-369-493-5013	Sequence 5013, Ap
28	86.5	10.2	443	9	US-09-815-242-14032	Sequence 14032, A
29	86.5	10.2	732	9	US-09-874-923-18	Sequence 18, Appl
30	86.5	10.2	732	10	US-09-991-496-18	Sequence 18, Appl
31	86.5	10.2	806	12	US-10-369-493-5292	Sequence 5292, Ap
32	86.5	10.2	891	12	US-10-369-493-10918	Sequence 10918, A
33	85.5	10.1	204	12	US-10-029-386-32082	Sequence 32082, A
34	85.5	10.1	734	12	US-10-104-047-2272	Sequence 2272, Ap
35	85.5	10.1	752	12	US-10-369-493-22302	Sequence 22302, A
36	85.5	10.1	884	12	US-10-032-585-7212	Sequence 7212, Ap
37	85.5	10.1	1007	12	US-10-369-493-22375	Sequence 22375, A
38	85	10.0	208	8	US-08-781-986A-5200	Sequence 5200, Ap
39	85	10.0	259	10	US-09-764-864-1313	Sequence 1313, Ap
40	85	10.0	636	9	US-09-815-242-5838	Sequence 5838, Ap
41	85	10.0	636	9	US-09-815-242-13008	Sequence 13008, A
42	84.5	10.0	390	10	US-09-738-626-3899	Sequence 3899, Ap
43	84.5	10.0	434	12	US-10-032-585-7146	Sequence 7146, Ap
44	84	9.9	932	9	US-09-815-242-5578	Sequence 5578, Ap
45	84	9.9	932	9	US-09-815-242-12438	Sequence 12438, A

ALIGNMENTS

RESULT 1  
US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score	848	DB	14	Length	773
Best Local Similarity	100.0%	Pred. No.	2.9e-68				
Matches	164	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	TTVKETILNDTGEVSELKP	RVTVTTIQNGKMSSTIVSEEDFTLPVYKGELEKGYFDG	60			
Db	610	TTVKETILNDTGEVSELKP	RVTVTTIQNGKMSSTIVSEEDFTLPVYKGELEKGYFDG	669			
Qy	61	WEISGFGGKDGAGVYINLSKDTFIKPVFKIEEKEEENKPTFDVSKKONPQVNHSQLN	120				
Db	670	WEISGFGGKDGAGVYINLSKDTFIKPVFKIEEKEEENKPTFDVSKKONPQVNHSQLN	729				
Qy	121	ESHRKEDLQREHSQKSDSKOVATATVLDKNNISSKSTNNPNK	164				
Db	730	ESHRKEDLQREHSQKSDSKOVATATVLDKNNISSKSTNNPNK	773				

```
RESULT 2
US-09-769-744A-28
; Sequence 28, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/F21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR FILING DATE: 1998-07-27
; PRIOR FILING DATE: 1998-07-27
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match          100.0%; Score 848; DB 12; Length 2119;
Best Local Similarity 100.0%; Pred. No. 1e-67; Indels 0; Gaps 0;
Matches 164; Conservative 0; Mismatches 0;

QY 1 TTWKEFILNDTCEVSELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQFDG 60
Db 1922 TTWKEFILNDTCEVSELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQFDG 1981

QY 61 WEISGEGKADAGYVNLKDTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLN 120
Db 1982 WEISGEGKADAGYVNLKDTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLN 2041

QY 121 ESHRKEDLOREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 164
Db 2042 ESHRKEDLOREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 2085

RESULT 3
US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
```

```
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68

Query Match          72.5%; Score 615; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.2e-48; Indels 0; Gaps 0;
Matches 117; Conservative 0; Mismatches 0;

QY 48 YKGELEKGYQFDGWEISGEGKADAGYVNLKDTFIKPVFKKIEKKEENKPTFDVSK 107
Db 1 YKGELEKGYQFDGWEISGEGKADAGYVNLKDTFIKPVFKKIEKKEENKPTFDVSK 60

QY 108 KDNQVNHSQLNESHKEDLOREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 164
Db 61 KDNQVNHSQLNESHKEDLOREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 117

RESULT 4
US-09-820-843A-107
; Sequence 107, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|3845248
US-09-820-843A-107

Query Match          12.5%; Score 106; DB 11; Length 665;
Best Local Similarity 24.3%; Pred. No. 0.33; Indels 56; Gaps 10;
Matches 42; Conservative 35; Mismatches 40;

QY 10 KDTGEVSELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQFDGWEI--SGF 66
Db 127 EEKNKINDLHRQNELNLSGK-----NEQDI-----NKNEKGKQ----DISNSNA 169

QY 67 EGKADAGYVNLKDTFIKPVFKKIEKKE-----BENKPTFD-----VSKKKNP 112
Db 170 ENKGD-----VKEGVKELEKKEEKESKSDHKVEENKSDHKVEENKSDDH 217

QY 113 QVNHSQLNESHKEDLOREHSQKSDTKDVTATVLDKNNISSKSTTNNPK 164
Db 218 KVEENKSDHKIEEVKVEEBHEDEE-----DKKKEKSENKKNKDNK 261

RESULT 5
US-10-381-596A-2
```

```
; Sequence 2, Application US/10381596A
; Publication No. US20040014178A1
; GENERAL INFORMATION:
; APPLICANT: Biostapro AB
; TITLE OF INVENTION: von Willebrand factor-binding proteins from
; FILE REFERENCE: 110059600
; CURRENT APPLICATION NUMBER: US/10/381,596A
; CURRENT FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: SE 0003573-3
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2060
; TYPE: PRT
; ORGANISM: Staphylococcus lugdunensis
US-10-381-596A-2

Query Match      11.4%; Score 96.5; DB 12; Length 2060;
Best Local Similarity 25.7%; Pred. No. 10;
Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;

Qy 14 EVSELKPHRVTTIQ-NKEMSSTIVSEEDFILPVYKGELEKGYQPDG--WEISGFEGKK 70
Db 1914 DIDLK---ITIVDTNGRE-----IVPSRKGLPP-EQFIGDQWQYTGHK--- 1955
Qy 71 DAGVINLSKDTFFKVPKIEE-----KKEENKPTFDVSKKONQVNHSQLNESH 123
Db 1956 -----IEKDGTITYIKKVENAVPAKQLKTHN--TQSEDFKHTPVQKQOLVKYHN 2006
Qy 124 RKE--DLQREHSOKSDSTKDVATVLDKNNI 153
Db 2007 VKQRSIEKSEHTDMHVELPETGETANNGI 2038

RESULT 6
US-10-172-502-10
; Sequence 10, Application US/10172502
; Publication No. US20030185833A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES...
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-10

Query Match      11.0%; Score 93.5; DB 12; Length 654;
Best Local Similarity 21.9%; Pred. No. 4.3;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

Qy 1 TTVKEFILNKDTGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 56
Db 336 SAITEFQNVQTNKWTDLQTKYVYVESVENESMMDTFVKH-----PIKTGMLNGKY 390
Qy 57 QF-----DGWEISGFEGKK-----D 71
Db 391 VMVETNDNDYKDFWVEQQRVTSKAKNNRTTIIPYVEGKTYDAIVKVVHKTIDYD 450
Qy 72 AGYVINL-SKDTFFKVPKKEEENKPTFDV-----SKKDNQVNHSQLNESH 125
Db 451 GQYHVRIVDKFAFTKANTDKSNKEQQDNSAKKEATPATPSKPTSPVPEKESQKDSQKD 510
Qy 126 EDLQ----REHSOKSDSTKDVATVLDKNNISSKSTTNNPNK 164

; Sequence 2, Application US/10381596A
; Publication No. US20040014178A1
; GENERAL INFORMATION:
; APPLICANT: Biostapro AB
; TITLE OF INVENTION: von Willebrand factor-binding proteins from
; FILE REFERENCE: 110059600
; CURRENT APPLICATION NUMBER: US/10/381,596A
; CURRENT FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: SE 0003573-3
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2060
; TYPE: PRT
; ORGANISM: Staphylococcus lugdunensis
US-10-381-596A-2

Query Match      11.4%; Score 96.5; DB 12; Length 2060;
Best Local Similarity 25.7%; Pred. No. 10;
Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;

Qy 14 EVSELKPHRVTTIQ-NKEMSSTIVSEEDFILPVYKGELEKGYQPDG--WEISGFEGKK 70
Db 1914 DIDLK---ITIVDTNGRE-----IVPSRKGLPP-EQFIGDQWQYTGHK--- 1955
Qy 71 DAGVINLSKDTFFKVPKIEE-----KKEENKPTFDVSKKONQVNHSQLNESH 123
Db 1956 -----IEKDGTITYIKKVENAVPAKQLKTHN--TQSEDFKHTPVQKQOLVKYHN 2006
Qy 124 RKE--DLQREHSOKSDSTKDVATVLDKNNI 153
Db 2007 VKQRSIEKSEHTDMHVELPETGETANNGI 2038

RESULT 6
US-10-172-502-10
; Sequence 10, Application US/10172502
; Publication No. US20030185833A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES...
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-10

Query Match      11.0%; Score 93.5; DB 12; Length 654;
Best Local Similarity 21.9%; Pred. No. 4.3;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

Qy 1 TTVKEFILNKDTGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 56
Db 336 SAITEFQNVQTNKWTDLQTKYVYVESVENESMMDTFVKH-----PIKTGMLNGKY 390
Qy 57 QF-----DGWEISGFEGKK-----D 71
Db 391 VMVETNDNDYKDFWVEQQRVTSKAKNNRTTIIPYVEGKTYDAIVKVVHKTIDYD 450
Qy 72 AGYVINL-SKDTFFKVPKKEEENKPTFDV-----SKKDNQVNHSQLNESH 125
Db 451 GQYHVRIVDKFAFTKANTDKSNKEQQDNSAKKEATPATPSKPTSPVPEKESQKDSQKD 510
Qy 126 EDLQ----REHSOKSDSTKDVATVLDKNNISSKSTTNNPNK 164

; Sequence 22285, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22285
; LENGTH: 1875
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22285

Query Match      10.8%; Score 92; DB 12; Length 1875;
Best Local Similarity 25.0%; Pred. No. 23;
Matches 48; Conservative 36; Mismatches 62; Indels 46; Gaps 10;

Qy 7 ILNKDTGVSSELKPHRVTVTIQNGKMSSTIVSEEDF-----ILPVYKGE---LEKGYOF 58
Db 660 LLNK---EIQLYDSKSDISIKLGRKSSRIAEERFALLNTLDLTKAENDQLKRPDY 716
Qy 59 DGWEISGFEGKKDA-----GYVINLSKDTFFKVPKIEEK-----KKEENK 100
Db 717 ----LQNTILKQDSKTHETLNEYVYVCKSLSVETELLNLKEQKLRVHLEKNLQELNK 772
Qy 101 PTFDVSKKONQVNHSQLNESH-KEDL---QREHSOKSDSTKDVTA-----TVLDKN 151
Db 773 ----LSPEKDSLRTIMVTQLQTKQEREDLLBETRKCKQKKTIDEALSELKETSQKDH 828
Qy 152 NISSKSTTNNPN 163
Db 829 HIKOLEEDNNSN 840

RESULT 8
US-10-032-585-7046
; Sequence 7046, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7046
; LENGTH: 635
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7046

Query Match      10.7%; Score 90.5; DB 12; Length 635;
Best Local Similarity 22.2%; Pred. No. 7.8;
Matches 38; Conservative 40; Mismatches 70; Indels 23; Gaps 8;
```





Db 502 KEFVRHKEKIQAKEA-VKENLKFSDSVKSTFRHFKDTTKNIFDE 547

## RESULT 12

US-09-827-664-4

; Sequence 4, Application US/09827664

; Patent No. US20020150956A1

; GENERAL INFORMATION:

; APPLICANT: Lawlor, Elizabeth J.

; TITLE OF INVENTION: No. US20020150956A1el tig

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESS: Dechert Price & Rhoads

; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

; CITY: Philadelphia

; STATE: PA

; COUNTRY: US

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/827,664

; FILING DATE: 06-Apr-2001

; CLASSIFICATION: <unknown>

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 09/464,483

; FILING DATE: <unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Falk, Stephen T

; REGISTRATION NUMBER: 36,795

; REFERENCE/DOCKET NUMBER: GMI0085

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-994-2488

; TELEFAX: 215-994-2222

; TELEX: <unknown>

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 402 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-827-664-4

Query Match 10.6%; Score 89.5; DB 10; Length 402;  
Best Local Similarity 22.1%; Pred. No. 5, 4;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

Qy 11 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK-----54

Db 53 DETDIKPAQEVSVTQIEKGKDFTEATVVEPEVKLGDKYGLIEKQETSLSDLEQE 112

Qy 55 -----GYQFDGWELSG--FEGKKDAGVYVNLKDTFTKP 86

Db 113 AIDHSLGHLAEMVVKEDGVWNGDTVTNIDFSG-SVDGEFEGGQAGYDLEIGSGFTI-P 170

Qy 87 VFK-----KTEEKE-----EE--NKPTFDVS---KKKDNPOVNHSQLNE- 121

Db 171 GFEEQLEGMKVDDEKDVVTFPEYHAELAGKEATFTKYNKIKFKEVPBELTDEIANEL 230

Qy 122 -----SHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTN 160

Db 231 DAEANTVDEYKENLKRKLAEQKTAENV-----EKEEAITKATDN 271

## RESULT 13

US-10-289-762-509

; Sequence 509, Application US/10289762

; Publication No. US20040006218A1

; GENERAL INFORMATION:

; APPLICANT: Griffois, R.

; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve

; TITLE OF INVENTION: and treatment of infection

; FILE REFERENCE: 9710-003-999

; CURRENT APPLICATION NUMBER: US/10/289,762

; CURRENT FILING DATE: 2003-03-27

; NUMBER OF SEQ ID NOS: 6849

; SEQ ID NO 509

; LENGTH: 511

; TYPE: PRT

; ORGANISM: Chlamydia pneumoniae

US-10-289-762-509

Query Match 10.6%; Score 89.5; DB 12; Length 511;

Best Local Similarity 24.5%; Pred. No. 7, 3;

Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

Qy 84 IKPVFKKIEEKKKEENKPTFD-----VSKKDNPOVNHSQLNESHRK 125

Db 95 VGVFKTTPQARPEVSSPRLPESHVHGQRLPGLSGFRDRIQKRSENPADLGKKRSYS 154

Qy 126 EDLQREHSQKSDTKDVTATVLDKNNISSKSTT 159

Db 155 GDLRVGHDSNEDSTEDSR---EGGEPSSKSS 185

## RESULT 14

US-10-032-585-7407

; Sequence 7407, Application US/10032585

; Publication No. US20030180953A1

; GENERAL INFORMATION:

; APPLICANT: Terry, Roemer D.

; APPLICANT: Bo, Jiang

; APPLICANT: Charles, Boone

; APPLICANT: Howard, Bussey

; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery

; FILE REFERENCE: 10182-005-999

; CURRENT APPLICATION NUMBER: US/10/032,585

; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 8000

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 7407

; LENGTH: 516

; TYPE: PRT

; ORGANISM: Candida albicans

US-10-032-585-7407

Query Match 10.6%; Score 89.5; DB 12; Length 516;

Best Local Similarity 23.3%; Pred. No. 7, 4;

Matches 37; Conservative 27; Mismatches 58; Indels 37; Gaps 6;

Qy 10 KDTGEVSELKPHRVTVT--IQNGKMSSTIVSEEDFILPVYKGELEKGQF-----58

Db 369 KNKGRIISRYLANKCSIASRIDYSEPTAFGE-----ILKKQVEDLKFYDTGSAPMK 422

Qy 59 -----DGWEISGFEGKDKAGVYVNLKDTFTKIPVKPIEKK-----BENKPTF 103

Db 423 NSDAIKAAALANGQDLAGAEQKQDMV---SDEQVKKKKKKKKKKKKKKKKKK 479

Qy 104 DVSKKDNPOVNHSQLNESHRKEDLQREHSQKSDTKD 142

Db 480 DKKEKKKKKKKKKK--DKKRKSDGGETPKKKKKKKSKD 516

## RESULT 15

US-09-827-664-2

; Sequence 2, Application US/09827664

; Patent No. US20020150956A1

; GENERAL INFORMATION:

; APPLICANT: Lawlor, Elizabeth J.

; TITLE OF INVENTION: No. US20020150956A1el tig

NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: PatsEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/827,664  
FILING DATE: 06-Apr-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/464,483  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Falk, Stephen T  
REGISTRATION NUMBER: 36,795  
REFERENCE/DOCKET NUMBER: GM10085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2488  
TELEFAX: 215-994-2222  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-827-664-2

Query Match 10.6%; Score 89.5; DB 10; Length 529;  
Best Local Similarity 22.1%; Pred. No. 7.6;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;  
QY 11 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 54  
Db 81 DETDIRKPAQEVSVTVQIEKGDIFFEATVPEVKGIDYKGLTEKQETLSDELQ 140  
QY 55 -----GYQFDGWEISG--FEGKQDAGYVINLSKDTFIKP 86  
Db 141 AIDHSILGLAEMVVKEDGVVNGDVTNIDFSG-SVDGEFEGGQAGYDLEIGSGSFI-P 198  
QY 87 VPK-----KIBEKE-----EE--NKPTFDVS-----KKONPQVNHSQLNE- 121  
Db 199 GFEEQLGKMKVDEKDVVTFPPEVHAEELAGKEATFKTVNEIKPKVEPPELTDEIANEL 258  
QY 122 -----SHRKDLQREHSQKSDSTKQVDTATVLDKNNISSKSTTN 160  
Db 259 DAEANTVDYKENLRKRLAEQKATDAENV-----EKKEAITRATDN 299

RESULT 16  
US-10-032-585-7829  
Sequence 7829, Application US/10032585  
Publication No. US20030180953A1  
GENERAL INFORMATION:  
APPLICANT: Terry, Roemer D.  
APPLICANT: Bo, Jiaog  
APPLICANT: Charles, Boone  
APPLICANT: Howard, Bussey  
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
FILE REFERENCE: 10182-005-999  
CURRENT APPLICATION NUMBER: US/10/032,585  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 8000  
SOFTWARE: PatentIn version 3.1

SEQ ID NO 7829  
LENGTH: 225  
TYPE: PRT  
ORGANISM: Candida albicans  
US-10-032-585-7829  
Query Match 10.4%; Score 88; DB 12; Length 225;  
Best Local Similarity 28.0%; Pred. No. 3.5;  
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;  
QY 56 YQFDGWEISGPEGKDGAGVINLSKDTFIKPVFKK-----IEEKEEENKPTFDVSKKK-- 109  
Db 80 YDDDDDEFESESSNGAAKELNLSQAIKEWKQRDLIEERERKLNKSKKEIEIEKAKS 139  
QY 110 --DNPQVNHSQLNESHKEDLQREH--SQKSDSTKQVDTATVLDKNN 152  
Db 140 TIDDFYENYNSKRDNHQKEILLSEKQFKISKRDDFLK--RGTILDRVN 184  
RESULT 17  
US-10-144-194A-106  
Sequence 106, Application US/10144194A  
Publication No. US20030215809A1  
GENERAL INFORMATION:  
APPLICANT: OriGene Technologies Inc  
TITLE OF INVENTION: Regulated Breast Cancer Genes  
FILE REFERENCE: 3U 103 R1  
CURRENT APPLICATION NUMBER: US/10/144,194A  
CURRENT FILING DATE: 2002-06-12  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 106  
LENGTH: 1298  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-144-194A-106  
Query Match 10.4%; Score 88; DB 12; Length 1298;  
Best Local Similarity 25.9%; Pred. No. 33;  
Matches 45; Conservative 23; Mismatches 60; Indels 46; Gaps 8;  
QY 10 KDTGSVSELKPHRVTVTQNGKMSSTIVSEEDFILPVYKGLEKGYQFDGWEISGFEK 69  
Db 944 KPSG-IASLQREPLLVNLSLVNENSGFTNIEIHSESYNKGIEISSGRKDNAAISGHSVE 1002  
QY 70 KDAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKK-----DNPQVNHSQL 118  
Db 1003 ADP-----KEVEE--EERHMPK--RKRKQHYLSSEDEPDNDPDVLDNR 1041  
QY 119 LNESHR-----KEDLQ--BEHSQKSDSTKQVDTATVL-DKNNISSKSTT 159  
Db 1042 IETAQRCQPETEPHPTKEENSRDLBELPKTSSETNSTTTSRVMEEKDEYSSETT 1095  
RESULT 18  
US-10-144-194A-62  
Sequence 62, Application US/10144194A  
Publication No. US20030215809A1  
GENERAL INFORMATION:  
APPLICANT: OriGene Technologies Inc  
TITLE OF INVENTION: Regulated Breast Cancer Genes  
FILE REFERENCE: 3U 103 R1  
CURRENT APPLICATION NUMBER: US/10/144,194A  
CURRENT FILING DATE: 2002-06-12  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 62  
LENGTH: 3051  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-144-194A-62  
Query Match 10.4%; Score 88; DB 12; Length 3051;

```
Best Local Similarity 25.9%; Pred. No. 96;
Matches 45; Conservative 23; Mismatches 60; Indels 46; Gaps 8;

Qy 10 KOTGEVSELPKRVTVTTIQNGKEMSTIVSEDFILPVYKGELEKGYQFDCGWEISGPEGK 69
Db 2697 KPSG-IAELQREPLLVNLSNVNENSGFTNEEIIHSESYNKGESISGRKNAEALSHGSHVE 2755
Qy 70 KDAGYVINLSKDTFKPVFKKIEBKKEENKPTFDVSKK-----DNPQVNHQS 118
Db 2756 ADP-----KEVEE--EERHMPK---RKRKQHYLSSEDEPDNDPDLDSR 2794
Qy 119 LNESHR-----KEDLQ--BEHSQKSDSTKQVTVATVL-DKNNISKSTT 159
Db 2795 IETAQRCQPETEPDHTKEENSRDLLELPKTSSETNSTTSRVMEBKDEYSSETT 2848

RESULT 19
US-10-369-493-2016
; Sequence 2016, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2016
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-2016

Query Match 10.3%; Score 87.5; DB 12; Length 655;
Best Local Similarity 20.5%; Pred. No. 15;
Matches 34; Conservative 36; Mismatches 57; Indels 39; Gaps 7;

Qy 6 FILNKOTGEVSELPKRVTVTTIQNGKEMSTIVSEDFILPVYKGELEKGYQF 58
Db 209 YIINKRTNSIK-----RSVSRTLRKGK-----TDSILPVYQSELKPFPRPSDDDDYKF 255
Qy 59 DGWEISGF--EGKKGAGYVINLSKDTFKPVFKKIEBKKEE-----ENKPTFDVSK 107
Db 256 TNIEDNKVREGR-----VHVSKESTADSTQKLGKEQKVIQSHLRHDDNNSTFRPHR 309
Qy 108 KKNPQV--NHSQNLNESHKEDLQREHSQKSDSTKQVTVATVLQKN 151
Db 310 LAPAPATKNHDSKTKWKHEDLLELKNDDNDDNEIMKMKTVDAIDVN 355

RESULT 20
US-09-815-242-5898
; Sequence 5898, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A

US-09-815-242-5898
; Sequence 5898, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5898
; LENGTH: 1349
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5898

Query Match 10.3%; Score 87.5; DB 9; Length 1349;
Best Local Similarity 25.0%; Pred. No. 38;
Matches 41; Conservative 23; Mismatches 73; Indels 27; Gaps 5;

Qy 9 NKDTGEVSELPKRVTVTTIQNGKEMSTIVSEDFILPVYKGELEKGYQFDCGWEISGF- 66
Db 804 NKDGKQDSTKSGISGVTVTLKN-----ENGEVLQTTKTDKQGYQFTGLENGTYK 853
Qy 67 -EGKKGAGYVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKDNPNQVNHQSOLNESHK 125
Db 854 VEFETPSGYTPT-----QVSGTDEGDSNCTSTTGVIKDKNDTIDSGFYKPTYNL 905
Qy 126 EDLQREHSQKSDSTKD-----VTATVLDKNNISKSTNNPN 163
Db 906 GDYVWEDTNKNGVQDKDEKISGVTVTLKDNENDKVLKTVTTDEN 949

RESULT 21
US-09-815-242-13137
; Sequence 13137, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A

US-09-815-242-13137
; Sequence 13137, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```

```
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13137
; LENGTH: 1349
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-13137

Query Match          10.3%; Score 87.5; DB 9; Length 1349;
Best Local Similarity 25.0%; Pred. No. 38;
Matches 41; Conservative 23; Mismatches 73; Indels 27; Gaps 5;

QY 9 NKTGEVSELKP-HRVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGP- 66
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :
Db 804 NDKGQDSTKRGISGVTVTLKN-----ENGEVLQTTKDKDKGYQFTGLENGTYK 853
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :

QY 67 -SGKXDAAGVVLNLSKDTFIKPKIEKKEENKPTFDVSKKKNPOVNHSQLNESHK 125
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :
Db 854 VEFETPSGYTPT-----QVSGTDEGIDNSGTSTTGVIKDKNDTIDSGFYKPTYNL 905
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :

QY 126 EDLQREEHQSQSDTKD-----VTATVLDKNNISSKSTNNPN 163
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :
Db 906 GDYVWEDTNKNGVQDKDEKGISSVTVTLKDENDKVLKTVTTDEN 949
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :

RESULT 22
US-10-032-585-7800
; Sequence 7800, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7800
; LENGTH: 1363
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7800

Query Match          10.3%; Score 87.5; DB 12; Length 1363;
Best Local Similarity 25.0%; Pred. No. 39;
Matches 41; Conservative 28; Mismatches 48; Indels 47; Gaps 10;

QY 10 KDTGEVSELKPHRVTVIQNG----KEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISG 65
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :
Db 569 KGIGDFHLKPHKVLQSSSSGDDVKTESLNVSKD-----TKG-----LKG 610
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :

QY 66 FPGKXDAAGVVLNLSKDT-----FIKPVFKIEKKEENKPTFDVSKKKNPOVNHSQLN 120
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :
Db 611 TDGRK---VVIDLYRTTPRIDBIEGHF-KLDDDHQETSYP-----HGEALIRHEAVN 659
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :

QY 121 ESHRKE--DLQREHSQSDSKDTVTATVLDKNNISSKSTNNPN 162
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :
Db 660 EWMRRKAVLFRKETELEKEGK-----LLDKN--ODQDQENKP 696
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :

RESULT 23
US-09-839-996-5
; Sequence 5, Application US/09839996
; Publication No. US20030009010A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; NUMBER OF SEQUENCES: 9

Query Match          10.3%; Score 87.5; DB 15; Length 1702;
Best Local Similarity 27.2%; Pred. No. 51;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/839,996
; APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-839-996-5

Query Match          10.3%; Score 87.5; DB 11; Length 1702;
Best Local Similarity 27.2%; Pred. No. 51;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 76 INLSKDTFIKPVFKIEKKEENKPTFDVSKKKNPOVNHSQLNESHKEDLQREHS- 134
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :
Db 1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRSI 1355
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :

QY 135 --QKSDTKDVTATVLDKNNISSKSTNNPNK 164
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :
Db 1356 SQPQTSAAETTAASTDDETTIADNSKRSKPNR 1387
   ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | :

RESULT 24
US-10-080-505-5
; Sequence 5, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-5

Query Match          10.3%; Score 87.5; DB 15; Length 1702;
Best Local Similarity 27.2%; Pred. No. 51;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;
```

QY 76 INLSKDTFTFKPVFKKIEKKBEENKPTFDVSKKKONPQNHSQLNESHKEDLQREHSH- 134  
||| : : : ||| : : : ||| : : : ||| : : :  
Db 1296 INTGSATAITETAESKDQPQTETAASTEDASQHKANTVADNSVANSSBPCKSRRRRSI 1355  
QY 135 --QKSDSTKDVATVLDKKNISSKSTTNNPNK 164  
: : : ||| : : : ||| : : : ||| : : :  
Db 1356 SQPQETSABETTAASTDETTIADNSKRSPNR 1387

RESULT 25  
US-10-369-493-16456  
; Sequence 16456, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 16456  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Bacillus thuringiensis  
US-10-369-493-16456

Query Match 10.3%; Score 87; DB 12; Length 323;  
Best Local Similarity 22.1%; Pred. No. 6.9;  
Matches 44; Conservative 33; Mismatches 84; Indels 38; Gaps 6;

QY 3 VKFIFLNKDTGVESELKPHRVTV--TIQNGKEMSSSTIVSEEDFILPVVK-----GELEK 54  
||| : : : ||| : : : ||| : : : ||| : : :  
Db 103 VKEHKQNID-NKWEIKEHKQTVDEKVNMKHQENIDQKVELKEVKQVDDECLAE LK 161  
||| : : : ||| : : : ||| : : : ||| : : :  
QY 55 GYDFDGWISGF-----EGKKDAGVVINLSKDTFI-----KP 86  
: : : ||| : : : ||| : : :  
Db 162 AKQTAEDKLAE LKENKPNGTNTLEELKTKIGNLDSLANSLELAKQDVKNKLAVLQEARQD 221  
QY 87 VFPKIKIEKEEENKPTFDVSKKKONPQNHSQLNESHK-EDLQREHSHSQKSDSTKDVTA 145  
||| : : : ||| : : : ||| : : : ||| : : :  
Db 222 LINKINEIKQSQTVDLSKKQDDLDIKINDFKITEKIDDKLAELHTTKONVDNKINE 281  
||| : : : ||| : : : ||| : : : ||| : : :  
QY 146 TVLDDKNISSKSTTNNPNK 164  
: : : ||| : : : ||| : : :  
Db 282 VSQSKOTQADNGTYKNNAK 300

RESULT 26  
US-10-369-493-21889  
; Sequence 21889, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374

```
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14032
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Salmonella typhi
; US-09-815-242-14032

Query Match      10.2%; Score 86.5; DB 9; Length 443;
Best Local Similarity 23.5%; Pred. No. 11;
Matches 43; Conservative 21; Mismatches 62; Indels 57; Gaps 8;

QY      16 SELKPHRVTVTTONKEMSTIVSE-----EDFILPVYKGELEKGYQ 57
Db      13 SRFTSLRTHITMNLTELKNTVPSELITLGSWGLENLARMKQDIIIFAILKQHAQSGED 72
QY      58 F-----DQWEISGEGKDGAGYVNLKSDTFIKP-----VFKKIEK 94
Db      73 IFDGVLEIQDGF---GFLRSADSSYLAG-PDDIYVSPSQIRRFNLRTGDTISGKIRPP 128
QY      95 KEE-----NKPTFDVSKKDNQVNHSQLNESHKREDIQRHEHSQKSDSTKDVAT 146
Db      129 KEGERYFALLKVNVDYKPNARN-KILFENLTPLHANSRLRME---RNGNSTEDLTAR 184
QY      147 VLD 149
Db      185 VLD 187

RESULT 29
US-09-874-923-18
; Sequence 18, Application US/09874923
; Patent No. US20020081320A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Coler, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C8
; CURRENT APPLICATION NUMBER: US/09/991,496
; CURRENT FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 732
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-991-496-18

Query Match      10.2%; Score 86.5; DB 10; Length 732;
Best Local Similarity 23.6%; Pred. No. 22;
Matches 39; Conservative 29; Mismatches 58; Indels 39; Gaps 8;

QY      6 FILNKDTGEVSELKPHRVTVTTONKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISG 65
Db      170 FTVRTDTGE-----PMGRGTVKVLHLKEDQTEYLEERRI-----KEIVKHSQFIGYPITL 220
QY      66 F-EGKKGAGYVNLKSDTFIKPVFKKIEEKEENKPTFDVSKKDNQVNHSQLNESHHR 124
Db      221 FVEKERDK-----EVSDD-----EAEKEDKEEKEEKESEDKPEI-----258
QY      125 KEDLQREHSQKSD-----STKDVATVLDKNNISSKST--TNNPN 163
Db      259 -EDVGSDEDEKKGDKGKKKKKIKKIKYIDKSELNKTPIWTRNPD 302

RESULT 30
US-09-991-496-18
; Sequence 18, Application US/09991496
; Patent No. US20020169285A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Coler, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C9
; CURRENT APPLICATION NUMBER: US/09/991,496
; CURRENT FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 732
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-991-496-18

Query Match      10.2%; Score 86.5; DB 10; Length 732;
Best Local Similarity 23.6%; Pred. No. 22;
Matches 39; Conservative 29; Mismatches 58; Indels 39; Gaps 8;

QY      6 FILNKDTGEVSELKPHRVTVTTONKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISG 65
Db      170 FTVRTDTGE-----PMGRGTVKVLHLKEDQTEYLEERRI-----KEIVKHSQFIGYPITL 220
QY      66 F-EGKKGAGYVNLKSDTFIKPVFKKIEEKEENKPTFDVSKKDNQVNHSQLNESHHR 124
Db      221 FVEKERDK-----EVSDD-----EAEKEDKEEKEEKESEDKPEI-----258
QY      125 KEDLQREHSQKSD-----STKDVATVLDKNNISSKST--TNNPN 163
Db      259 -EDVGSDEDEKKGDKGKKKKKIKKIKYIDKSELNKTPIWTRNPD 302

RESULT 31
US-10-369-493-5292
```







```
;
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781.986A
; FILING DATE:
;
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: P8248PP
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 5200:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 208 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-781-986A-5200
;
; Query Match 10.0%; Score 85; DB 8; Length 208;
; Best Local Similarity 21.7%; Pred. No. 6;
; Matches 38; Conservative 22; Mismatches 89; Indels 26; Gaps 5;
;
; Qy 9 NKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGPEG 68
; Db 27 NSSKNDTNKTKTQTDNTQSTKQMTQPEAEDIVRNDYKARGVNEYQTILNYKTNLRS 86
; Qy 69 KKDAGYVNLKSDTFIKP-----VFKIEEKGEEENKPTFDVSKK---K 109
; Db 87 NEHEYVEHLVRDAVGTPLKCAIVNRNGIINIFDDMSKDEE-----FEAFKRSPPK 142
; Qy 110 DNPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 164
; Db 143 YNPGWNHDETGG-ESEDIQHHDIDNNKAIQNDIPDQKVDDKN--DKNAVNKEEK 194
;
; RESULT 39
; US-09-764-864-1313
; Sequence 1313, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1313
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (251)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
;
; ADDRESS: (255)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (259)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-764-864-1313
;
; Query Match 10.0%; Score 85; DB 10; Length 259;
; Best Local Similarity 23.8%; Pred. No. 7.9;
; Matches 36; Conservative 30; Mismatches 51; Indels 34; Gaps 8;
;
; Qy 17 ELKPHRVTVTIQ-----NGKEM--SSTIVSEEDFILPVYKGELEKGYQ-FDGEWEISGF 66
; Db 93 EMKNDRIKVSLSMKVNVNQGTGKOLDPNNVIIQEE-----RRRRSFQDYTGOKIT-L 143
; Qy 67 EGK-----KDAGYVNLKSDTFIKPVPKJ-----EKKKEENKPTFDVSKKDNQVNVN 115
; Db 144 EAVLNTTCKGCGCKGHGFAKCFMQPGGTGYKSLIPDEEEKEEAKSAEFE-----KPDPTRN 199
; Qy 116 HSQNLNESHKEDLQREHSQKSDSTKDVATAT 146
; Db 200 PSRRKKEKKKKKRRDRKSSDSDSDSESMT 230
;
; RESULT 40
; US-09-815-242-5838
; Sequence 5838, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 60/191,078
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5838
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5838
;
; Query Match 10.0%; Score 85; DB 9; Length 636;
; Best Local Similarity 22.0%; Pred. No. 25;
; Matches 35; Conservative 35; Mismatches 61; Indels 28; Gaps 7;
;
; Qy 7 ILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGF 66
; Db 27 IVTKDYSGKSVQVNSKNGKTLIDSRVLSALYLYEDYI--IYAGLTNKKYEY-----GD 78
; Qy 67 EGKDGAGYVNLKSDTFIKPVPK-----IEBKKEE-ENKPTFDVSKKDNQVNVNSQLN- 120
```

Db 79 NIYKEA-----KRLLEKVLREDQYLLERKKSQYEDYKQWYANYKENPRTLKMANF 131  
QY 121 ESHRKEDLOREHHSQKSDTKDV-----TATVLDKNN 152  
Db 132 HKNLEELSMKEYNELQDALKRALDDFHREVKDKDKNS 170

Search completed: February 10, 2004, 11:23:56  
Job time : 28.9011 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 11.1197 Seconds  
(without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799  
Sequence: 1 DTGVSSELKPHRVTTIQNG.....ATVLDKNNISKSTNNPNK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_76:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	799	100.0	2140	2 F95074	serine proteinase,
2	796	99.6	2144	2 A97942	metalloproteinase,
3	110	13.8	558	2 T18467	hypothetical prote
4	108.5	13.6	1038	2 JCS497	claustrin - chicke
5	105	13.1	665	2 B71609	hypothetical prote
6	103.5	13.0	325	2 T18283	hypothetical prote
7	100	12.5	312	2 G81339	probable membrane
8	99	12.4	211	2 T25911	hypothetical prote
9	97.5	12.2	614	2 A84152	hypothetical prote
10	97.5	12.2	1345	2 S46817	hypothetical prote
11	97	12.1	385	2 T20410	hypothetical prote
12	97	12.1	988	2 T14188	hypothetical prote
13	96.5	12.1	540	2 D86432	hypothetical prote
14	96	12.0	535	2 T37189	hypothetical prote
15	96	12.0	2500	2 G71609	hypothetical prote
16	95.5	12.0	644	2 T47835	hypothetical prote
17	95	11.9	348	2 I37271	cylicin II - human
18	95	11.9	622	2 A90570	lipoprotein (impor
19	95	11.9	1397	2 T10466	DNA topoisomerase
20	94.5	11.8	3724	2 T18427	hypothetical prote
21	94	11.8	210	2 T28771	hypothetical prote
22	93.5	11.7	456	2 T05612	hypothetical prote
23	93	11.6	219	2 B72291	hypothetical prote
24	91	11.4	253	2 T32879	hypothetical prote
25	90.5	11.3	629	2 G36542	hypothetical prote
26	90.5	11.3	670	2 T28391	ORF MSV230 hypothe
27	90.5	11.3	867	2 T27136	hypothetical prote
28	90.5	11.3	871	2 T27135	hypothetical prote
29	90.5	11.3	1332	2 S41552	probable transcrip

30	90.5	11.3	2401	2 T28676	rhoptyr protein -
31	90	11.3	645	2 E89883	conserved hypothet
32	90	11.3	1202	1 S05362	probable DNA-direc
33	89.5	11.2	433	2 A89951	trigger factor [im
34	89.5	11.2	508	2 B81594	hypothetical prote
35	89.5	11.2	508	2 C72074	hypothetical prote
36	89.5	11.2	508	2 E86549	hypothetical prote
37	89.5	11.2	700	2 S67610	probable membrane
38	89.5	11.2	1016	2 T19006	ankyrin related pr
39	89	11.1	528	2 E96795	unknown protein P2
40	89	11.1	1888	2 T39009	hypothetical prote
41	88.5	11.1	301	2 T33088	hypothetical prote
42	88.5	11.1	371	2 A71683	hypothetical prote
43	88.5	11.1	443	2 S66040	serine-type D-Ala-
44	88.5	11.1	762	2 G88436	protein T04A8.13 [
45	88.5	11.1	791	2 T24435	hypothetical prote

ALIGNMENTS

RESULT 1

F95074

serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: F95074

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,

Science 293, 498-506, 2001

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: F95074

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2140 <CUR>

A:Cross-references: KB:AE005672; PIDN:AAK74791.1; PID:g14972117; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0641

Query Match 100.0%; Score 799; DB 2; Length 2140;

Best Local Similarity 100.0%; Pred. No. 7.2e-52;

Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTGVSSELKPHRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEGKK 60

Db 1953 DTGVSSELKPHRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEGKK 2012

Qy 61 DAGVINLSKDTFKPVFKIEEKEENKPTFDVSKKNDPQVNHSLNESHKEDLQR 120

Db 2013 DAGVINLSKDTFKPVFKIEEKEENKPTFDVSKKNDPQVNHSLNESHKEDLQR 2072

Qy 121 BEHSQKSDSTKDTATVLDKNNISSKSTNNPNK 154

Db 2073 BEHSQKSDSTKDTATVLDKNNISSKSTNNPNK 2106

RESULT 2

A97942

metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C:Accession: A97942

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A97942  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99365.1; PID:g15458138; GSPDB:GN00174  
C:Genetics:  
A:Gene: ptaA  
C:Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 796; DB 2; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 1.2e-51;  
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGVESELKPHRVTTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEK 60  
Db 1957 DTGVESELKPHRVTTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEK 2016

QY 61 DAGYVNLKSDTFTDVKPKIEKKEENKPTFDVSKKDNPOVNHSQLNESHKEDLQR 120  
Db 2017 DAGYVNLKSDTFTDVKPKIEKKEENKPTFDVSKKDNPOVNHSQLNESHKEDLQR 2076

QY 121 EHSQKSDSTKDTATVLDKNNISSKSTNNPNK 154  
Db 2077 EHSQKSDSTKDTATVLDKNNISSKSTNNPNK 2110

RESULT 3  
T18467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2002  
C:Accession: T18467  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: T18467  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-558 <LAW>  
A:Cross-references: EMBL:AL008970; NID:e1407852; PIDN:CAA15610.1  
C:Genetics:  
A:Map position: 3  
A:Introns: 84/1; 160/1  
A:Note: C0465c

Query Match 13.8%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 0.69;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 27 IVSEEDFILPVY-----KGELEKGYQFDGWEISGFEK-----DAGYVNLKSDTFTKPV 77  
Db 60 ILGFEDDILVEYCISLQKQKEK--KADGEEDKYLNAKKLKNLTGFIGNKKSDIFIEL 117

QY 78 FPKI--EKKKE-----ENKPTFDVSK-KKDNPOVNHSQLNE-----SHRK 115  
Db 118 LELLINEEKKEHIAIDLNENK-TNDIKVQENENINENYNNENKDNKDKHVSQHN 176

QY 116 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISSKSTTN 150  
Db 177 EHNNNNVLKKEKYTDIQDKKHKRSLSQKSDSYKKRPFNKRKTSIER-SLSNKRIDE 235

QY 151 NPNK 154  
Db 236 KTNK 239

RESULT 4  
JCS497  
claustrin - chicken  
N:Alternate names: keratan sulfate proteoglycan  
C:Species: Gallus gallus (chicken)  
C:Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 21-Jul-2000  
C:Accession: JCS497; PC4334; S37561

R:Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A:Title: Claustarin, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JCS497; MUID:94157526; PMID:7906711  
A:Accession: JCS497  
A:Molecule type: mRNA  
A:Residues: 1-1038 <BUR1>  
A:Cross-references: EMBL:X67778; NID:g406318; PIDN:CAA47988.1; PID:g406319  
A:Accession: PC4334  
A:Molecule type: protein  
A:Residues: 79-83;299-412;485-502 <BUR2>  
A:Experimental source: brain  
C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervous  
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F:267-270/Region: cell attachment (R-G-D) motif  
F:112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 13.6%; Score 108.5; DB 2; Length 1038;  
Best Local Similarity 28.6%; Pred. No. 1.8;  
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

QY 21 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKSDTFTKPVFK- 79  
Db 595 KPETKTIVAEKDV-----TTKEQLGKSETSEKQASEKQDVKPKVTKESVKKEVKA 646

QY 80 KIEKKEEENKPTFDVSKKDNPOVNHSQLNESHKEDLQRE-----EHSQKSDSTKDV 133  
Db 647 KPKEKKDEKPKPKVSKKEKPLI---KSEKPKKEDIKKEVKKEKKEKKEAKKEV 702

RESULT 5  
B71609  
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: B71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Perte, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: B71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-665 <GAR>  
A:Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AACT1925.1; PID:g384524  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0680w

Query Match 13.1%; Score 105; DB 2; Length 665;  
Best Local Similarity 24.4%; Pred. No. 2;  
Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;

QY 1 DTGVESELKPHRVTTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEI--SGFE 57  
Db 128 EKNKINKSLHRLQNELNLQSGK-----NEQDI-----NKNKRGKQ-----DISNSNAE 170

QY 58 GKGDAGYVNLKSDTFTKPVFKKIEKKE-----EENKPTFD-----VSKKDNPO 103  
Db 171 NKKD-----VKEGVKELEKKEKKEIKSDHKVKEENKSDDDHKVKEENKSDDDHK 218

QY 104 VNHSQLNESHKEDLQRE-EHSQKSDSTKDTATVLDKNNISSKSTNNPNK 154  
Db 219 VEENKSDDDHKIEEVKVEEHEDEEE-----DKKEKSENKNKNDENK 261

RESULT 6  
T18283  
hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999





Db 2281 LEESE---KSDKRD-----DKNDNTRKNNLDNK 2308

RESULT 14

T37189

hypothetical protein C02H7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000

C:Accession: T37189

R:Leimbach, D.; Minx, M.

A:Submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid C02H7.

A:Reference number: Z20523

A:Accession: T37189

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-535 <LEI>

A:Cross-references: EMBL:U49945; PIDN:AAC47924.1; GSPDB:GNO00029; CESP:C02H7.1

A:Experimental source: strain Bristol N2; clone C02H7

C:Genetics:

A:Gene: CESP:C02H7.1

A:Map position: X

A:Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 12.0%; Score 96; DB 2; Length 535;

Best Local Similarity 21.3%; Pred. No. 7.4;

Matches 35; Conservative 28; Mismatches 61; Indels 40; Gaps 5;

Qy 1 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 60

Db 77 DQSLKVNKAAKII---SGDAET-----NKMQLMGTNATSFNSRN 116

Qy 61 DAGYVINLSKDTIFKVPFKIEEKEENKPTFDVSKKDNQPNVHSQLNESHRKEDLQR 120

Db 117 GTG-----BEKKKKVKEDKKGDEEKST---TKRSSKTHEEKEKSKKSAEE 166

Qy 121 EHSQKSSTK-----DVTATVLDKNNISSKSTNNPNK 154

Db 167 KEKKKSSSRERHKSSDRSRSEKSEKSKSEKKEKSTTDEKPK 210

RESULT 15

G71609

hypothetical protein PFB0650w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C>Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C:Accession: G71609

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: G71609

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2500 <GAR>

A:Cross-references: GB:AE001408; GB:AE001362; NID:g3845238; PIDN:AAC71919.1; PID:g384524

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0650w

Query Match 12.0%; Score 96; DB 2; Length 2500;

Best Local Similarity 26.1%; Pred. No. 42;

Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;

Qy 9 KPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINL 68

Db 2173 KPYKIT---ENKK-----NEGNILKKYSTENEKKNYDKEQNECILDKDTQCNVNT 2223

Qy 69 -----SKDTFKPVFKIEEKEENKPTFDVSKKDNQPNVHSQLNE-----SHRKED 117

Db 2224 KEKNLNDNKKSPFNKVKLEEEKSDKRD---DKNDNTRKNNLDNKKSFPSNIVK 2280

Qy 118 LQREHSHQKSDSKDVTATVLDKNNISSKSTNNPNK 154

Db 119 LQREHSHQKSDSKDVTATVLDKNNISSKSTNNPNK 154

Db 2281 LEESE---KSDKRD-----DKNDNTRKNNLDNK 2308

RESULT 16

T47835

hypothetical protein T209.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T47835

R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Stert, W.; Holland, R.; Weichselgartner, M.;

A:Submitted to the Protein Sequence Database, February 2000

A:Reference number: Z24475

A:Accession: T47835

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-644 <NVA>

A:Cross-references: EMBL:AL138658

A:Experimental source: cultivar Columbia; BAC clone T209

C:Genetics:

A:Map position: 3

A:Introns: 158/2; 329/3

A:Note: T209.90

Query Match 12.0%; Score 95.5; DB 2; Length 644;

Best Local Similarity 23.4%; Pred. No. 10;

Matches 34; Conservative 22; Mismatches 58; Indels 31; Gaps 4;

Qy 18 QNGKEMSTIVSE---DFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTF 73

Db 519 ENSKTEKTVADKKKSVADFLKRIKKNSPQKG-----ETTSKNQKNDGNV----- 565

Qy 74 IKPVFKIEEKEENKPTFDVSKKDNQPNVHSQLNESHRKEDLQREHSHQKSDS----- 129

Db 566 -----KKENDHQKSDGNVKKENSKVYKPRELASSSTGKKKVEVNNNNSSSKRKQ 615

Qy 130 TKQVTVATVLDKNNISSKSTNNPNK 154

Db 616 TKETAEVATGKRGESGKDDKQPRK 640

RESULT 17

I37271

cylicin II - human

C:Species: Homo sapiens (man)

C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000

C:Accession: I37271; S52774

R:Hees, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.

A:Exp. Cell Res. 218, 174-182, 1995

A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the

A:Reference number: I37271; MUID:95255491; PMID:7737358

A:Accession: I37271

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-348 <HES>

A:Cross-references: EMBL:Z46788; NID:g758586; PIDN:CAA86752.1; PID:g758587

Query Match 11.9%; Score 95; DB 2; Length 348;

Best Local Similarity 30.9%; Pred. No. 5.4;

Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

Qy 39 KGELEKGYQFDGWEISGFEKKDAGYVINLSKDTIFKPVFKIEEKEENKPTF---DV 95

Db 205 ESEGEK-----GTEKSKKGGKDS-----KKGKDSAIELQAVKADEKDEDGKDANKGDE 256

Qy 96 SK---KKNQPNVHSQLN-----ESHRKEDLQREHSHQKSDSKD---VTATVLDKNNI 143

Db 257 SKDAKDAKEIKGGKDKKKPSSTDSKDDVKE---SKDATKDAKKVAKDTEKESA 313

Qy 144 SSK 146

Db 314 DSK 316





C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
C:Accession: T05612  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohelsel, J.; Mewes, H.W.; Mayer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05612  
A:Molecule type: DNA  
A:Residues: 1-456 <BEV>  
A:Cross-references: EMBL:AL035394  
A:Experimental source: cultivar Columbia; BAC clone F9D16  
C:Genetics:  
A:Map position: 4  
A:Introns: 110/3; 247/2; 282/3; 304/3; 361/3; 390/2; 418/3  
A:Note: F9D16.270

Query Match 11.7%; Score 93.5; DB 2; Length 456;  
Best Local Similarity 25.0%; Pred. No. 9.5;  
Matches 39; Conservative 28; Mismatches 58; Indels 31; Gaps 8;

Qy 4 EVSELPK-HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDA 62  
Db 249 EKDPKPKHPSAFIVYANERAAALREENKSVVEAK-----ITGEENKNSDKKA 300  
Qy 63 GY--VINLSKDTFIKPVFKIEEKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120  
Db 301 PYEKVAKKNKETYLG-AMEEYKRTKEE-----ALSQKE-----EEELLKLHQEALQM 349  
Qy 121 EEHSOKSDTKDVTATVLDKNNISSKSTNN--PNK 154  
Db 350 LKKEKTDN-----LIKKEKATKKKKNENVDNK 378

RESULT 23  
B72291  
hypothetical protein - Thermotoga maritima (strain MSB)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: B72291  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: B72291  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <ARN>  
A:Cross-references: GB:AE001771; GB:AE000512; NID:g4981678; PIDN:AAD36218.1; PID:g498168  
A:Experimental source: strain MSB  
C:Genetics:  
A:Gene: TM1142

Query Match 11.6%; Score 93; DB 2; Length 219;  
Best Local Similarity 26.2%; Pred. No. 4.5;  
Matches 42; Conservative 28; Mismatches 50; Indels 40; Gaps 9;

Qy 8 LKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGY--FDGWEISG-----PEGK 59  
Db 64 LIPNVFVELVATDEKTLFAKE-----VLGEESVSRYDLFAGFGVGRGTPPTFFFPKGK 116  
Qy 60 KDAGVIN-LKDTFIKPVFKIEEKEENKPTFDVSKKDNPNQVNHSQLNESHKED- 117  
Db 117 EGLGLPGVQKDNFIK-ILKYVAQELKED-----PQTYLKKDDPPVGRPLIETP-KEDA 170  
Qy 118 ---LQREHSOKSDS-----TKDVTATVLDK 140  
Db 171 DFVLEKDNVAVKQVTPNVRDRRIYVTSDPDVAKTLQEK 210

RESULT 24  
TJ32879

hypothetical protein C17F3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32879  
R:Gattung, S.; Scheet, P.  
submitted to the EMBL Data Library, January 1998  
A:Description: The sequence of C. elegans cosmid C17F3.  
A:Reference number: Z21240  
A:Accession: T32879  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-253 <GAT>  
A:Cross-references: EMBL:AF043692; PIDN:AAB97531.1; GSPDB:GN00019; CESP:C17F3.3  
A:Experimental source: strain Bristol N2; clone C17F3  
C:Genetics:  
A:Gene: CESP:C17F3.3  
A:Map position: 1  
A:Introns: 41/1

Query Match 11.4%; Score 91; DB 2; Length 253;  
Best Local Similarity 32.6%; Pred. No. 7.6;  
Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;

Qy 76 PVFKIEEKE---EENKPTFDVSKK-----KDNQVNHSQLNESHK----- 114  
Db 61 PVAPKVEEKEEKEEKADDEKKTEEKDDKSKKTEEKDKISVKTKQETKSERKDKK 120  
Qy 115 ---KEDLQREHSOKSDSTKDVATVLDKNNISSK 146  
Db 121 DEKDEKKEGSEKSKDEEKKDEKDKDEKDK 155

RESULT 25  
G96542  
hypothetical protein F17J6.14 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: G96542  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.;  
anssen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G96542  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-629 <STO>  
A:Cross-references: GB:AB005173; NID:g11054631; PIDN:AAG27876.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F17J6.14  
A:Map position: 1

Query Match 11.3%; Score 90.5; DB 2; Length 629;  
Best Local Similarity 19.5%; Pred. No. 23;  
Matches 33; Conservative 37; Mismatches 66; Indels 33; Gaps 5;

Qy 10 PHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGEGKDGAVINLS 69  
Db 423 PHLTAKPTKDSAMEQTEAEADVAMNPIVEKAMSEMVEAGAAINPIVEADG-----A 476  
Qy 70 KDTFIKPVFKIEEKE---EENKPTFDVSKKDNPNQ---VNHSQLNESHK 116  
Db 477 MNPIVEKAMSOIVEADAAINQADVAMNPQATGNDDAESDDFSEPVSHS---ETLNPP 533  
Qy 117 DLQREHSOKSDSTKDVATVLDKNN-----ISSKSTNNPNK 154  
Db 534 ELEKKEVMRKDATERSVSADCCQDKNSKIIESSIQEELISASQTSNPTQ 582

## RESULT 26

T28391  
ORF MSV230 hypothetical protein - Melanoplus sanguinipes entomopoxvirus  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T28391  
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28391  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-670 <AFO>  
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97713.1; PID:g4049753  
C:Genetics:  
A:Note: MSV230

Query Match 11.3%; Score 90.5; DB 2; Length 670;  
Best Local Similarity 22.5%; Pred. No. 25;  
Matches 36; Conservative 34; Mismatches 69; Indels 21; Gaps 5;  
  
QY 15 VTIQKEMSSIT-----VSEDFILPVYKGELEKGY-----QFDGWEISGF 56  
Db 33 VNFEEKQIITLLKFNFDKTEMCGVSEKVFQLINNKASAEKSDVDSSIDESQNSDS 92  
QY 57 EGKDGAGYVILNSKDTFIKPVFKIE-EKKEENKPTFDVSKKNDPQVNHSQL-NESHK 114  
Db 93 DSDSDSGVNIDESQNSDKVINKLENSQNSDKVINDESQNSDS-KVINKLENSQN 151  
QY 115 KEDLQREHSQKSDSTKYVTATVLDKNNISSKSTTNNPK 154  
Db 152 SDSKVINDESQNSDKVINDESQNSDKVINDESQNSDK 191

## RESULT 27

T27136  
hypothetical protein Y53C12B.3a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27136  
R:Kershaw, J.; Lennard, N.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z20316  
A:Accession: T27136  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-867 <WIL>  
A:Cross-references: EMBL:Z99278; PIDN:CAB16493.1; GSPDB:GN000020; CESP:Y53C12B.3a  
A:Experimental source: clone Y53C12B  
C:Genetics:  
A:Map position: 2  
A:Introns: 100/3; 177/3; 218/1; 423/3; 714/2; 864/3

Query Match 11.3%; Score 90.5; DB 2; Length 867;  
Best Local Similarity 26.4%; Pred. No. 33;  
Matches 42; Conservative 27; Mismatches 57; Indels 33; Gaps 7;  
  
QY 4 EVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELE-----KGYQFDGWEISGFEGK 59  
Db 672 EISEENPK--TDDIQSKDDVTS-----KSELHCYRCHEHYQLPAEEVSSHNR 716  
QY 60 KDAGYVILNSKDTFIKPVFKIE-----EKKEENKPTFDVSKKNDPQVNHSQLNESHK 115  
Db 717 KNG-----DLWCHEHM-KKIKGCHCEATGEGHPLICPKKKEERVAKSR--ESSQK 766  
QY 116 EDLQREHSQKSDSTKYVTATVLDKNNISSKSTTNNPK 154  
Db 767 PIDPQISDDQDDTVPDQIVQEQDNQSHKSRHNSNR 805

## RESULT 28

T27135  
hypothetical protein Y53C12B.3b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27135  
R:Kershaw, J.; Lennard, N.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z20316  
A:Accession: T27135  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-871 <WIL>  
A:Cross-references: EMBL:Z99278; PIDN:CAB16492.1; GSPDB:GN000020; CESP:Y53C12B.3b  
A:Experimental source: clone Y53C12B  
C:Genetics:  
A:Gene: CESP:Y53C12B.3b  
A:Map position: 2  
A:Introns: 100/3; 177/3; 218/1; 423/3; 717/2; 867/3

Query Match 11.3%; Score 90.5; DB 2; Length 871;  
Best Local Similarity 26.4%; Pred. No. 33;  
Matches 42; Conservative 27; Mismatches 57; Indels 33; Gaps 7;  
  
QY 4 EVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELE-----KGYQFDGWEISGFEGK 59  
Db 675 EISEENPK--TDDIQSKDDVTS-----KSELHCYRCHEHYQLPAEEVSSHNR 719  
QY 60 KDAGYVILNSKDTFIKPVFKIE-----EKKEENKPTFDVSKKNDPQVNHSQLNESHK 115  
Db 720 KNG-----DLWCHEHM-KKIKGCHCEATGEGHPLICPKKKEERVAKSR--ESSQK 769  
QY 116 EDLQREHSQKSDSTKYVTATVLDKNNISSKSTTNNPK 154  
Db 770 PIDPQISDDQDDTVPDQIVQEQDNQSHKSRHNSNR 808

## RESULT 29

S41552  
probable transcription factor SPT7 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YBR0739; protein YBR081C  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-Jan-1994 #sequence\_revision 09-Sep-1994 #text\_change 20-Sep-1999  
A:Accession: S41552; S45946; S45948; S40800; S45478; S54985; S59716  
R:Gansheroff, L.; Dollard, C.; Fan, P.; Winston, F.  
submitted to the EMBL Data Library, July 1993  
A:Reference number: S41552  
A:Accession: S41552  
A:Molecule type: DNA  
A:Residues: 1-1332 <GAN>  
A:Cross-references: EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:g349190  
R:Steensma, H.Y.; van der Aart, Q.J.M.  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S45932  
A:Accession: S45946  
A:Molecule type: DNA  
A:Residues: 1-1332 <STE>  
A:Cross-references: EMBL:Z35950; NID:g536341; PIDN:CAAB5026.1; PID:g536342; MIPS:YBR081C  
R:Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.; Urrestazu, A.; Vissers, S.  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S45893  
A:Accession: S45948  
A:Molecule type: DNA  
A:Residues: 1-835 <AND>  
A:Cross-references: EMBL:Z35950; MIPS:YBR081C  
R:Haynes, S.R.; Dollard, C.; Winston, F.; Beck, S.; Trowsdale, J.; Dawid, I.B.  
Nucleic Acids Res. 20, 2603, 1992  
A:Title: The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins  
A:Reference number: S40800; MUID:92285152; PMID:1350857  
A:Accession: S40800  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA

```

A:Accession: A45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: GB:M34281

Query Match      11.3%; Score 90.5; DB 2; Length 2401;
Best Local Similarity 27.4%; Pred. No. 1e+02;
Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;

QY      16  TIQNGKEMSTI-----VSEEDFILPVYKGELEKGY---QFDGWEISGFEKKDA 62
      QY      16  TIQNGKEMSTI-----VSEEDFILPVYKGELEKGY---QFDGWEISGFEKKDA 62
      DB      16  TIQNGKEMSTI-----VSEEDFILPVYKGELEKGY---QFDGWEISGFEKKDA 62
      QY      63  GYVINLSKOTFIKPVFKKBEKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQREE 122
      QY      63  GYVINLSKOTFIKPVFKKBEKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQREE 122
      DB      63  GYVINLSKOTFIKPVFKKBEKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQREE 122
      QY      983  IY--NILKLNKIKKIIDKVKEYTDEIEK-----NNKINAELSSEKIIITQLKNSSLKE 1035
      QY      983  IY--NILKLNKIKKIIDKVKEYTDEIEK-----NNKINAELSSEKIIITQLKNSSLKE 1035
      DB      983  IY--NILKLNKIKKIIDKVKEYTDEIEK-----NNKINAELSSEKIIITQLKNSSLKE 1035
      QY      123  HSQKSOSTKD-----VTATVLDKKNISS 145
      QY      123  HSQKSOSTKD-----VTATVLDKKNISS 145
      DB      123  HSQKSOSTKD-----VTATVLDKKNISS 145
      QY      1036  CQSKIKSTIDDNVYSECIRKTNLKYIVNEKKNINT 1072
      QY      1036  CQSKIKSTIDDNVYSECIRKTNLKYIVNEKKNINT 1072
      DB      1036  CQSKIKSTIDDNVYSECIRKTNLKYIVNEKKNINT 1072

RESULT 31
E89883
conserved hypothetical protein SA0976 [imported] - Staphylococcus aureus (strain
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: E89883
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui,
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekim
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: E89883
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-645 <KUR>
A:Cross-references: GB:BA000018; PID:g13700929; PIDN:BAB42225.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA0976

Query Match      11.3%; Score 90; DB 2; Length 645;
Best Local Similarity 21.4%; Pred. No. 26;
Matches 45; Conservative 33; Mismatches 67; Indels 66; Gaps 10;

QY      4  EVSELKPHRVTV--TTQNGKEMSTIVSEEDFILPVYKGELE-KGYOF-----DGWEIS 54
      QY      4  EVSELKPHRVTV--TTQNGKEMSTIVSEEDFILPVYKGELE-KGYOF-----DGWEIS 54
      DB      4  EVSELKPHRVTV--TTQNGKEMSTIVSEEDFILPVYKGELE-KGYOF-----DGWEIS 54
      QY      341  KMTDLQDTKYVWYVESVENNESMMDTFVKH-----PIKTGMLNGKKYVMVNETTNDYDKDF 395
      QY      341  KMTDLQDTKYVWYVESVENNESMMDTFVKH-----PIKTGMLNGKKYVMVNETTNDYDKDF 395
      DB      341  KMTDLQDTKYVWYVESVENNESMMDTFVKH-----PIKTGMLNGKKYVMVNETTNDYDKDF 395
      QY      55  GFEGKK-----DAGYVNL--SKDTFI 74
      QY      55  GFEGKK-----DAGYVNL--SKDTFI 74
      DB      55  GFEGKK-----DAGYVNL--SKDTFI 74
      QY      396  MVSGQRVTTISKDAKNTRTIIIPYVEGKTYDAIVKVVHVKTIDYDQCYHVRIVDKRAFT 455
      QY      396  MVSGQRVTTISKDAKNTRTIIIPYVEGKTYDAIVKVVHVKTIDYDQCYHVRIVDKRAFT 455
      DB      396  MVSGQRVTTISKDAKNTRTIIIPYVEGKTYDAIVKVVHVKTIDYDQCYHVRIVDKRAFT 455
      QY      75  KPVEFKTEEEKKEENKPTFDV-----SKKQNPQVNHSQLNESHKEDLQ-----REEHSQ 125
      QY      75  KPVEFKTEEEKKEENKPTFDV-----SKKQNPQVNHSQLNESHKEDLQ-----REEHSQ 125
      DB      75  KPVEFKTEEEKKEENKPTFDV-----SKKQNPQVNHSQLNESHKEDLQ-----REEHSQ 125
      QY      456  KANTDKSNKEQQDNSAKKATPATSKPTFSPVEKESQKQDSQKQDNKQLPSVEKENDA 515
      QY      456  KANTDKSNKEQQDNSAKKATPATSKPTFSPVEKESQKQDSQKQDNKQLPSVEKENDA 515
      DB      456  KANTDKSNKEQQDNSAKKATPATSKPTFSPVEKESQKQDSQKQDNKQLPSVEKENDA 515
      QY      126  KSDSTKDVT--ATVLDKKNISSKSTTNNPNK 154
      QY      126  KSDSTKDVT--ATVLDKKNISSKSTTNNPNK 154
      DB      126  KSDSTKDVT--ATVLDKKNISSKSTTNNPNK 154
      QY      516  SSESQKDKTPATKPTKGEVSSSTT--PTK 543
      QY      516  SSESQKDKTPATKPTKGEVSSSTT--PTK 543
      DB      516  SSESQKDKTPATKPTKGEVSSSTT--PTK 543

RESULT 32
S05362
probable DNA-directed DNA polymerase (EC 2.7.7.7) - fungus (Ascobolus immersus)
C:Species: mitochondrion Ascobolus immersus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S05362
R:Kempken, F.; Meinhardt, F.; Esser, K.
Mol. Gen. Genet. 218, 523-530, 1989

```

A:Title: In organello replication and viral affinity of linear, extrachromosomal DNA of  
A:Reference number: S05362; MUID:90066356; PMID:2573821  
A:Accession: S05362  
A:Molecule type: DNA  
A:Residues: 1-1202 <KEM>  
A:Cross-references: EMBL:X15982; NID:g2933; PIDN:CAA34106.1; PID:g1370212  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC3  
C:Superfamily: Ascolobus probable DNA-directed DNA polymerase  
C:Keywords: DNA binding; mitochondrion; nucleotidyltransferase

Query Match 11.3%; Score 90; DB 1; Length 1202;  
Best Local Similarity 22.6%; Pred. No. 52;  
Matches 40; Conservative 35; Mismatches 74; Indels 28; Gaps 8;

QY 2 TGEVSELKPHRVTVTTONKEMSSSTI---VSEEDF--ILPVVYKGELEK-----GVQFQD 50  
DB 321 TGNVRSIGFNGVT--TLTDXETLIKILAFLEREDITHVMSYDEGIDSKFPKGLSDFD 379

QY 51 WEISGFEKKDAGYVNLKDTFIKPVFKKIE-----EKBEENK-PTFDVSKKDNPNQ 103  
DB 380 KPLKTIEGKYANTYTPFKKDIWVKDINKINFNGLDLPKTDLSKWPNLKLNKDKTSGE 439

QY 104 VHSQNLNESHKEDLOREHSQKSDTKQVTVATVLDKNNI-----SSKSTNNPN 153  
DB 440 IRTMIKNKNQSYDI--IGHMIINDGENVITFNRAVDNSIIKFTVTDSMGNTNDPN 494

RESULT 33  
A89951  
trigger factor [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: A89951  
R:Kuroda M.; Ohta T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.; Ma, A.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: A89951  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-433 <KUR>  
A:Cross-references: GB:BA000018; PID:g13701472; PIDN:BAB42766.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: tlg  
C:Superfamily: trigger factor; BKBP-type peptidylprolyl isomerase homology

Query Match 11.2%; Score 89.5; DB 2; Length 433;  
Best Local Similarity 22.1%; Pred. No. 18;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 1 DTGEVSELKPHRVTVT-IONGKE--MSSTIVSEEDFILPVYKG-ELEK----- 44  
DB 81 DETDIKPAQVSVSTQIEKQDFIEATVTVPEVKLGDKYKGLTEKQETELSDDELQ 140

QY 45 -----GYQFDGWEISG--FEKKDAGYVNLKDTFIK 76  
DB 141 AIDHSLGHIAEMVVKEDGVVNGDVTNIDFSG--SVDDGEFGGQARGYDLEIFGSGSFI-P 198

QY 77 VFK-----KIEKKE-----EE--NKPTFDVS-----KKKDNFPQVNHSLNF- 111  
DB 199 GFEEQLGKMKVDEKDVVTFPEYHAEELAGKEATFKTKVNEIKFKEVPELTDEIANEL 258

QY 112 -----SHRKEDLOREHSQKSDTKQVTVATVLDKNNISSKSTTN 150  
DB 259 DAEANTVDEYKENLKRRLAEQKATDAENV-----EKEEAITKATDN 299

RESULT 34

B81594  
hypothetical protein CP0281 [imported] - Chlamydophila pneumoniae (strain AR39)  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
A:Accession: B81594  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, M.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255; PMID:10684935  
A:Accession: B81594  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <REA>  
A:Cross-references: GB:AB002189; GB:AE002161; NID:g7189205; PIDN:AAF38139.1; PID:g7189202  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CP0281

Query Match 11.2%; Score 89.5; DB 2; Length 508;  
Best Local Similarity 24.5%; Pred. No. 21;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 74 IKPVFKTTEKKEENKPTFD-----VSKKDNFPQVNHSQLNESHK 115  
DB 92 VGVFKFTQARPEVSSPRLFSHVQHGQRLFCLEGFRDRIQKRSNPEDLGKMKRSYSD 151

QY 116 EDLOREHSQKSDTKQVTVATVLDKNNISSKSTT 149  
DB 152 GDLDRVGHDSNEDSTEDSRs---EGGEPSKSSS 182

RESULT 35  
C72074  
hypothetical protein - Chlamydophila pneumoniae (strain CWL029)  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
A:Accession: C72074  
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606; PMID:10192388  
A:Accession: C72074  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <ARN>  
A:Cross-references: GB:AE001632; GB:AE001363; PIDN:AAI8613.1; PID:g4376755; PID:g4376752  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: Cpn0473

Query Match 11.2%; Score 89.5; DB 2; Length 508;  
Best Local Similarity 24.5%; Pred. No. 21;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 74 IKPVFKTTEKKEENKPTFD-----VSKKDNFPQVNHSQLNESHK 115  
DB 92 VGVFKFTQARPEVSSPRLFSHVQHGQRLFCLEGFRDRIQKRSNPEDLGKMKRSYSD 151

QY 116 EDLOREHSQKSDTKQVTVATVLDKNNISSKSTT 149  
DB 152 GDLDRVGHDSNEDSTEDSRs---EGGEPSKSSS 182

RESULT 36  
E86549  
hypothetical protein CPj0473 [imported] - Chlamydophila pneumoniae (strain J138)  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
A:Accession: E86549  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A;Accession: T19006

```
Db      257 RYSDVKIPVYDSELTSEPSKNGESNTNEKEKDISTENHLESTALNIQQSDSTPTP 316
QY      133 -----VTATVLDKQNISSKSTNN 151
          |||  :  :  :
Db      317 MEEDVVTVETKTETSEDMLLSQN 340

RESULT 40
T39009
hypothetical protein SPAC6B12.02c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T39009
R:Gentiles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A:Reference number: Z21815
A:Accession: T39009
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1888 <GEN>
A:Cross-references: EMBL:Z98531; PIDN:CAB11064.1; GSPDB:GN00066; SPDB:SPAC6B12.02c
A:Experimental source: strain 972h-; cosmid c6B12
C:Genetics:
A:Gene: SPDB.SPAC6B12.02c
A:Map position: 1
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC6B12.02c

Query Match      11.1%; Score 89; DB 2; Length 1888;
Best Local Similarity 23.9%; Pred. No. 1e+02;
Matches 37; Conservative 29; Mismatches 59; Indels 30; Gaps 7;

QY      24 SSTIVSEEDF-----ILPVY---KGELEKGYQFDGWEISGFE-----GKKDAG 63
          ||: |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      390 SSSLTSENPFQLNVAANAVSTIPVYRTTKTKMKKN-RFKYVEVEKLPDLILESYGKKAPK 448

QY      64 YVINLSKDTFIKPVFKKIEEKEENKPTFDYKKDNQVNVHSQLNESHKEDLQREEH 123
          ::  ::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      449 FLRVFARSSHIP--KMIIRKQMDSKKYFDFKESDRQVIDQVLSDWYSGKHELQQOSH 506

QY      124 SQKSDS-TKDVATVLDKN-----NISSKSTNN 151
          |  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      507 SYKKPSDSKSVGGNIFSVNSKKHVNINAKTAANN 541

Search completed: February 10, 2004, 10:58:34
Job time : 12.1197 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 7.15925 Seconds  
(without alignments)  
1011.574 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799

Sequence: 1 DTGEVSELKPHRVTTIQNG.....ATVLDKNISSKSTNNPNK 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97.5	12.2	1345	1 YH00 YEAST	P38800 saccharomyc
2	95	11.9	348	1 CYL2 HUMAN	Q14093 homo sapien
3	95	11.9	1398	1 TOP2 PLAFK	P41001 plasmodium
4	92.5	11.6	2468	1 MAPB HUMAN	P46821 homo sapien
5	90.5	11.3	1332	1 SP7 YEAST	P35177 saccharomyc
6	90	11.3	1202	1 DPOM ASCIM	P22374 ascobolus i
7	89.5	11.2	433	1 TIG STAAH	Q99216 staphylococ
8	89	11.1	1888	1 YDT2 SCHPO	O14207 schizosacch
9	88.5	11.1	443	1 DACA BACSU	P08750 bacillus su
10	87.5	11.0	1702	1 IGA2 HAEIN	P45384 haemophilus
11	87	10.9	893	1 GYA2 STAEP	P54112 staphylococ
12	87	10.9	2464	1 MAPB MOUSE	P14873 mus musculu
13	86.5	10.8	1006	1 RAT1 YEAST	Q02792 saccharomyc
14	86.5	10.8	1875	1 MLP1 YEAST	Q02455 saccharomyc
15	86	10.8	443	1 GLNA PYRKO	O08467 pyrococcus
16	86	10.8	5596	1 MDN1 HUMAN	Q98222 homo sapien
17	85.5	10.7	914	1 PBPA BACSU	P39793 bacillus su
18	85.5	10.7	1007	1 RGAI YEAST	P39083 saccharomyc
19	85	10.6	720	1 IF2 STAEP	Q8C8t4 staphylococ
20	85	10.6	6632	1 UN89 CAEEL	O01761 caenorhabdi
21	84.5	10.6	1694	1 IGA0 HAEIN	P44969 haemophilus
22	84	10.5	778	1 YFK8 YEAST	P43610 saccharomyc
23	83.5	10.5	479	1 U2R1 HUMAN	Q15695 homo sapien
24	83.5	10.5	943	1 ARS2 DROME	Q9V9K7 drosophila
25	83	10.4	258	1 IF31 HUMAN	O75822 homo sapien
26	83	10.4	270	1 TONB HAEIN	P42872 haemophilus
27	83	10.4	688	1 LIP STAEP	Q02510 staphylococ
28	83	10.4	715	1 ADSV HUMAN	Q9Y6u3 homo sapien
29	83	10.4	847	1 DNLI CAEEL	Q24774 caenorhabdi
30	82.5	10.3	439	1 GLNA PYRAB	Q9UY99 pyrococcus
31	82.5	10.3	649	1 HS70 PYRSA	P37899 pyrenomonas
32	82.5	10.3	2459	1 MAPB RAT	P15205 rattus norv
33	82	10.3	279	1 YMB4 YEAST	P49957 saccharomyc

ALIGNMENTS

RESULT 1

ID	YH00 YEAST	STANDARD;	PRT;	1345 AA.
AC	P38800;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	Hypothetical 149.7 kDa protein in IREL-KSP1 intergenic region.			
GN	YHR080C.			
OS	Saccharomyces cerevisiae (Baker's Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288c / AB972;			
RX	MEDLINE=94378003; PubMed=8091229;			
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,			
RA	Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,			
RA	Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,			
RA	Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,			
RA	Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,			
RA	Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,			
RA	Vaudin M.;			
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome			
RT	VIII.";			
RL	Science 265:2077-2082(1994).			
CC	-I- SIMILARITY: TO YEAST YFL042C.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U10556; AAB68895.1; ..	
PIR; S46817; S46817	
SGD; S0001122; YHR080C.	
InterPro; IPR004182; GRAM_dom.	
Pfam; PF02893; GRAM; 1.	
SMART; SM00568; GRAM; 1.	
HyPOthetical protein; Transmembrane.	
TRANSMEM 1198 1218	POTENTIAL.
SEQUENCE 1345 AA; 149679 MW; 2FDADB94A686564C2 CRC64;	

Query Match	12.2%;	Score 97.5;	DB 1;	Length 1345;
Best Local Similarity	27.0%;	Pred. No. 7.9;		
Matches	33;	Conservative	18;	Mismatches 38;
				Indels 33;
				Gaps 6;
Qy	39	KGELEKGVQPCWISGEGKK-DAGVIVNLKSTFTFKVPFKKIEEKEENKPTFDYSK	97	
		::		
		::		
Db	1109	KGAIKEG-----SVGQKSVVDYMLSELDII-----		---SRAKSKPVKVMK 1149
Qy	98	KKDNPNVNSQLNESHKREDLQREHSQKSDTKDVTATVLD--KNNISSKSTTN--		---NP 152

```
Db 1150 SHDKHPFISKVE-----QKSSBSRSDDNKOILTHILDFVQNNFSSEIFMKNLLSP 1201
QY 153 NK 154
Db 1202 QK 1203

RESULT 2
CYL2_HUMAN
ID CYL2_HUMAN STANDARD; PRT; 348 AA.
AC Q14093;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cyclicin II (Multiple-band polypeptide II).
GN CYL2 OR CYL2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95355491; PubMed=7737358;
RA Hess H., Heid H., Zimbelmann R., Franke W.W.;
RT "The protein complexity of the cytoskeleton of bovine and human sperm
RT heads: the identification and characterization of cyclicin II.";
RL Exp. Cell Res. 218:174-182(1995).
CC -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -!- TISSUE SPECIFICITY: Testis.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z46788; CAA86752.1; -
CC PIR; I37271; I37271.
CC Genew; HGNC:2583; CYL2.
CC MIM; 604035; -
CC GO: GO:0005200; F:structural constituent of cytoskeleton; TAS.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 25 347 31 X 3 AA REPEATS OF K-K-X.
FT REPEAT 157 240 3 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 157 184 1.
FT REPEAT 185 212 2.
FT REPEAT 213 240 3.
SQ SEQUENCE 348 AA; 39079 MW; D86766599C1809E7 CRC64;

Query Match 11.9%; Score 95; DB 1; Length 348;
Best Local Similarity 30.9%; Pred. No. 2.8;
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 39 KGELEKGYQDFGWEISGFGKDKAGVINLSKDTFTFKVPFKIEEKKKEENKPTP-----DV 95
Db 205 ESRGEGK-----GTEKDSKKGKDS-----KKGKDSALQVADKEDGDKDKANGDE 256

QY 96 SK--KKDNPOVNSQLN-----ESHKEDLQREHSOKSDSTKD---VTATVLDDKNNI 143
Db 257 SKDAKKDAKEIKGKKKKKPSSTDSKDDVAKK---SKKATKDAKKVAKDTKEKSA 313

QY 144 SSK 146
Db 314 DSK 316

RESULT 3
```

```
TOP2_PLAFK
ID TOP2_PLAFK STANDARD; PRT; 1398 AA.
AC P41001;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94316496; PubMed=8041616;
RA Cheesman S., McAltese S., Goman M., Johnson D., Horrocks P.,
RA Ridley R.G., Kilbey B.J.;
RT "The gene encoding topoisomerase II from Plasmodium falciparum.";
RL Nucleic Acids Res. 22:2547-2551(1994).
CC -!- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMS
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X79345; -; NOT ANNOTATED CDS.
CC HSSP; P06786; 1BGM.
CC InterPro; IPR003594; ATPbind ATPase.
CC InterPro; IPR001957; CBFA_NFYB_topis.
CC InterPro; IPR001241; DNA_topoisoi.
CC InterPro; IPR002205; DNA_topoisoi.
CC Pfam; PF00204; DNA_gyraseB; 1.
CC Pfam; PF00521; DNA_topoisoi; 1.
CC Pfam; PF02518; HATPase_c; 1.
CC PRINTS; PR00615; CCAATSUBUNTA.
CC PRINTS; PR00418; TP12FAMILY.
CC ProDom; PD000742; DNA_topoisoi; 1.
CC SMART; SM00387; HATPase_c; 1.
CC SMART; SM00433; TOP2c; 1.
CC SMART; SM00434; TOP4c; 1.
CC PROSITE; PS00177; TOPOISOMERASE II; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
FT NP_BIND 144 149 ATP (POTENTIAL).
FT ACT_SITE 830 830 DNA CLEAVAGE (BY SIMILARITY).
FT DOMAIN 271 281 POLY-ASN.
FT DOMAIN 308 316 POLY-ASN.
FT DOMAIN 1089 1093 POLY-LYS.
FT DOMAIN 1227 1234 POLY-LYS.
SQ SEQUENCE 1398 AA; 161029 MW; BAAD7BEE88FE5BE9 CRC64;

Query Match 11.9%; Score 95; DB 1; Length 1398;
Best Local Similarity 23.8%; Pred. No. 13;
Matches 36; Conservative 36; Mismatches 36; Indels 26; Gaps 6;

QY 19 NGKMSSTIVSEEDFIL--PVYKGELEKGYQDFGWEISGFGKDKAGVINLSKDTFTFKP 76
Db 1133 NEETIAGITVKDYDYLMSPIFSLTLEK---VEDLLTQKERELEILRNITVETMLK 1189

QY 77 VPKIEE-----KKEENKPTFDVSKKDNPOVNSQLNESHKEDLQREHSQ 125
```



Db 1190 DIEKVEAIEFORNVLSNRRESNK--FKVARKQ-----GPSMMKKKKKKKKLSSDBESE 1242

Qy 126 ---KSDTKDVTATVLDKNNISSKSTNNPN 153

Db 1243 GGDTSDSSEFLVNLINKNTNKKTTTSSNN 1273

RESULT 4

MAPB HUMAN

ID MAPB HUMAN STANDARD; PRT; 2468 AA.

AC P46821;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain LC1].

DE LC1].

GN MAP1B.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxID=9606;

EN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RX MEDLINE=95104835; PubMed=7806212;

RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;

RT "Cloning of human microtubule-associated protein 1B and the identification of a related gene on chromosome 15.";

RL Genomics 22:273-280(1994).

CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN. PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN STABILIZING MICROTUBULES.

CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE WITH MAP1A AND MAP1B PROTEINS.

CC -!- DOMAIN: Has a highly basic region with many copies of the sequence KKEE and KKEI/V repeated but not at fixed intervals, which is responsible for the binding of MAP1B to microtubules.

CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated from MAP1B by proteolytic processing. It is free to associate with both MAP1A and MAP1B. It interacts with the amino-terminal region of MAP1B (By similarity).

CC -!- SIMILARITY: TO MAP1A.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; L06237; AAA18904.1; -

CC Genew; HGNC:6836; MAP1B.

CC MIM; 157129; -

CC GO; GO:0005875; C: microtubule associated complex; TAS.

CC InterPro; IPR000102; MAP1B neuraxin.

CC Pfam; PF00414; MAP1B neuraxin; 10.

CC PROSITE; PS00230; MAP1B NEURAXIN; 6.

CC Microtubules; Repeat; Phosphorylation.

CC CHAIN ? 2468

CC MAP1 LIGHT CHAIN LC1.

FT REPEAT 1878 1894

FT REPEAT 1895 1911

FT REPEAT 1912 1928

FT REPEAT 1929 1945

FT REPEAT 1946 1962

FT REPEAT 1963 1979

FT REPEAT 1997 2013

FT REPEAT 2014 2030

FT REPEAT 2031 2047

FT REPEAT 2048 2064

FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY KKEE AND KKEI/V REPEATS).

SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;

Query Match 11.6%; Score 92.5; DB 1; Length 2468;

Best Local Similarity 31.0%; Pred. No. 35;

Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Qy 68 LSKDTFIKPVFKKIEKKKEENKPTFDVSKKNQPNQVHSLQSHRKLQRE-----E 122

Db 638 VKKETVKP-----EDKKEKEKPKKEVAKEDKTPi---KKEKPKKEEVKKEVKEIK 689

Qy 123 HSKQSDSTKDV 133

Db 690 KEEKKEPKKEV 700

RESULT 5

SPT7 YEAST

ID SPT7 YEAST STANDARD; PRT; 1332 AA.

AC P35177;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Transcriptional activator SPT7.

GN SPT7 OR YBR081C OR YBR0739.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.

OC NCBI\_TaxID=4932;

EN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=95229044; PubMed=7713415;

RA Gansheroff L.J., Dollard C., Tan P., Winston F.;

RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein important for transcription in vivo.";

RL Genetics 139:523-536(1995).

CC [2]

CC SEQUENCE FROM N.A.

CC STRAIN=S288C;

CC MEDLINE=95076715; PubMed=7985423;

CC van der Aart Q.J.M., Barthe C., Doignon F., Aigle M., Crouzet M., Steensma H.V.;

CC "Sequence analysis of a 31 kb DNA fragment from the right arm of Saccharomyces cerevisiae chromosome II.";

CC Yeast 10:959-964(1994).

CC [3]

CC SEQUENCE OF 1-835 FROM N.A.

CC STRAIN=S288C;

CC Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A., Vissers S.;

CC Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.

CC [4]

CC SEQUENCE OF 463-523 FROM N.A.

CC MEDLINE=92285152; PubMed=1350857;

CC Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J., David I.B.;

CC "The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins.";

CC Nucleic Acids Res. 20:2603-2603(1992).

CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY OTHER GENES.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- SIMILARITY: Contains 1 bromodomain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC EMBL; L22537; AAC37424.1; -.
DR EMBL; X76294; CAA53940.1; -.
DR EMBL; Z35950; CAA85026.1; -.
DR EMBL; M87651; AAA35087.1; -.
DR PIR; S41552; S41552.
DR HSP; Q92831; I891.
DR TRANSFAC; T04835; -.
DR SGD; S000285; SPT7.
DR GO; GO:000124; C:SAGA complex; IDA.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS00633; BROMODOMAIN 1; 1.
DR PROSITE; PS00114; BROMODOMAIN 2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain.
FT DOMAIN 458..528 BROMODOMAIN.
SQ SEQUENCE 1332 AA; 152616 MW; 083B63624669244F CRC64;

Query Match 11.3%; Score 90.5; DB 1; Length 1332;
Best Local Similarity 24.5%; Pred. No. 26;
Matches 39; Conservative 26; Mismatches 57; Indels 37; Gaps 8;

QY 15 VTIQKEMSSITVSEEDFILPVYKGELEKGYQDFGWEISGEGKDGAGVYINLSKDTFI 74
DB 549 IITRNADLEKET---EDM-----EKDKDYELDEEEVAGSGRKG-----LNMGAHMLA 594
QY 75 K---PVFKKIEEKEEENKPTFD-----VSKKID-----NPQVNHLSNESH 113
DB 595 KENGKVSSEKSSKTVKDEAFTNDKLTSVIPEGEKDKTASSTVTVHENVKNEIKENG 654
QY 114 RKEDLQR-BEHSQKSDTKDVTATVLD-KNNISSKSTTN 150
DB 655 KNEEQDMVESSKTESDSSKDAAKKTDGLQDKTAEN 693

RESULT 6
DPOM_ASCIM STANDARD; PRT; 1202 AA.
AC P22374;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable DNA polymerase (EC 2.7.7.7).
OS Ascobolus immersus.
OG Mitochondrion.
OC Plasmid pA12.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;
OC Pezizales; Ascobolaceae; Ascobolus.
OX NCBI_TaxID=5191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2/I;
RX MEDLINE=90066356; PubMed=2573821;
RA Kempken F., Meinhardt F., Esser K.;
RT "In organello replication and viral affinity of linear,
RL extrachromosomal DNA of the ascomycete Ascobolus immersus.";
RL Mol. Gen. Genet. 218:523-530(1989).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N).
CC -!- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG, TO
CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; X15982; CAA34106.1; -.
DR PIR; S05362; S05362.
DR InterPro; IPR006172; DNA pol B.
DR InterPro; IPR004868; DNA pol B 2.
DR Pfam; PF03175; DNA pol B 2; 1.
DR SMART; SM00486; POLB; 1.
DR PROSITE; PS00116; DNA_POLYMERASE B; 1.
KW Transferrase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Plasmid; Mitochondrion.
SQ SEQUENCE 1202 AA; 138279 MW; 51D1FCBBDBF2CDE CRC64;

Query Match 11.3%; Score 90; DB 1; Length 1202;
Best Local Similarity 22.6%; Pred. No. 25;
Matches 40; Conservative 35; Mismatches 74; Indels 28; Gaps 8;

QY 2 TOEVELKPHRVTVTIQNGKMSSTI---VSEEDF--ILPVYKGELEK-----GYQFDG 50
DB 321 TGNVRSIGFQVNT-TLTQKTLIKTLAIFLEREDHTVMSYDEGDI DESKPKGSLSDPF 379
QY 51 WEISGEGKDGAGVYINLSKDTFIKPVFKKIE-----EKKEENK-PTFDVSKKDNQP 103
DB 380 RPLKTIETGKYANYTFPIKDIWVKDKINFNGLDKPTWDLKSWNLKLNKDKTSGE 439
QY 104 VNHSLNESHKEDLQREHSQKSDTKDVTATVLDKNNI-----SSKSTNNPN 153
DB 440 IRMTIKNNQSYDI--IGHMIINDGVNFTNRAVDNSIIKFTVTDMSGNTNDPN 494

RESULT 7
TIG_STAAM STANDARD; PRT; 433 AA.
ID TIG_STAAM
AC Q99TIG;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trigger factor (TF).
GN TIG OR SAV1675 OR SAI499 OR MM1619.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 196620;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RT Lancet 357:1225-1240(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MW2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
CC -!- FUNCTION: Involved in protein export. Acts as a chaperone by
CC maintaining the newly synthesized protein in an open conformation
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY. TIG SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AP003363; BAB57837.1; -  
 DR EMBL; AP003134; BAB42766.1; -  
 DR EMBL; AP004827; BAB95484.1; -  
 DR PIR; A89951; A89951.  
 DR HAMAP; MF\_00303; -; 1.  
 DR InterPro; IPR001179; FKBP\_PP1ASE.  
 DR InterPro; IPR005215; Trig\_fac.  
 DR Pfam; PF00254; FKBP; 1.  
 DR TIGRFAMs; TIGR00115; tig; 1.  
 DR PROSITE; PS00453; FKBP\_PP1ASE\_1; FALSE\_NEG.  
 DR PROSITE; PS00454; FKBP\_PP1ASE\_2; FALSE\_NEG.  
 DR PROSITE; PS00509; FKBP\_PP1ASE\_3; 1.  
 DR Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.  
 FT DOMAIN 163 248 PPIASE, FKBP-TYPE.  
 SQ SEQUENCE 433 AA; 48609 MW; 8865D9AF6A18C1E7 CRC64;

Query Match 11.2%; Score 89.5; DB 1; Length 433;  
 Best Local Similarity 22.1%; Pred. No. 9.1;  
 Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;  
 Yy 1 DTGVSSELKPHRVTVT-IQNGKE--MSTIVSEEDFILPVYKG-ELEK----- 44  
 Db 81 DETDIKVAQPEVSVTQIEKGKDFEATVVEPEVGLGDKYGLIEIKQETELSDDELQ 140  
 Yy 45 -----GYPDQWEISG--PEGKDGAGVIVNLSKDTIKP 76  
 Db 141 AIDSLGHLAMVVKEDGVVNGDTVVIDFSG-SVDGEFEGGQAGYDLIGSGSFI-P 198  
 Yy 77 VFK-----KLEEKKE-----EE--NRPTEFVS---KKDNPNQVNSHLSNE- 111  
 Db 199 GFEEQLGEMKVDDEKVVVTPPEYHAEELAGKATPKTKVNEIKFKEVPELTDIEAN 258  
 Yy 112 -----SHRKEDLQREHSQKSDSTKDTATVLDKNNISKSTTN 150  
 Db 259 DAEANTVDYKENLRKRLAEQKATDAENV-----EKEEAITKATDN 299

RESULT 8  
 YD2 SCHPO STANDARD; PRT; 1888 AA.  
 AC O14207;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C6B12.02c in chromosome I.  
 GN SPAC6B12.02C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouris J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Murgill K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Gailbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Sipakowski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z98531; CAB11064.1; -  
 DR PIR; T39009; T39009.  
 DR GeneDB Spombe; SPAC6B12.02c; -  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 697 717 POTENTIAL.  
 FT TRANSMEM 755 775 POTENTIAL.  
 FT TRANSMEM 866 886 POTENTIAL.  
 FT TRANSMEM 915 935 POTENTIAL.  
 FT TRANSMEM 977 997 POTENTIAL.  
 FT TRANSMEM 1033 1053 POTENTIAL.  
 FT TRANSMEM 1336 1356 POTENTIAL.  
 FT TRANSMEM 1645 1665 POTENTIAL.  
 FT DOMAIN 1662 1665 POLY-LEU.  
 SQ SEQUENCE 1888 AA; 217432 MW; 8AD3BBCE32397C29 CRC64;

Query Match 11.1%; Score 89; DB 1; Length 1888;  
 Best Local Similarity 23.9%; Pred. No. 48;  
 Matches 37; Conservative 29; Mismatches 59; Indels 30; Gaps 7;

Yy 24 SSTIVSEEDF-----ILPVY---KGELEKGYQFDGWEISGFE-----GKDKAG 63  
 Db 390 SSSLTSENPFQNLVNAANAVSTIPVYRTTKTKRKN-RFKYVEVEKLPDLILESCKKAPK 448  
 Yy 64 YVINLSKDTFIKVPFKIIEKKKEENKPTFDVSKKDNPNQVNSHLSHREKEDLQREH 123  
 Db 449 FLRVFARSSSHIP--KMIRKQMDSKKYFSFDKESDRQVIDQVLSDWYSGKHELQVQSH 506  
 Yy 124 SQKSDS-TKDVATVLDKN-----NISKSTTN 151  
 Db 507 SYKKPSDSKSVGGNIFSVNSKKGSHVNINAKTAANN 541

RESULT 9  
 DACA BACSU STANDARD; PRT; 443 AA.  
 AC P08750;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE D-alanyl-D-alanine carboxypeptidase precursor (EC 3.4.16.4) (DD-  
 DE peptidase) (DD-carboxypeptidase) (CPase) (PBPs).  
 GN DACA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;

RX MEDLINE=960511385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunat F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruechi C., Caldwel B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,  
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kashara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel J.C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler H., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
RN [3]  
RP SEQUENCE OF 32-102.  
RX MEDLINE=80182289; PubMed=6768745;  
RA Waxman D.J., Strominger J.L.;  
RT "Sequence of active site peptides from the penicillin-sensitive D-  
RT alanine carboxypeptidase of Bacillus subtilis. Mechanism of  
RT penicillin action and sequence homology to beta-lactamases.";  
RL J. Biol. Chem. 255:3964-3976(1980).  
RN [4]  
RP SEQUENCE OF 103-443 FROM N.A.  
RX MEDLINE=86250602; PubMed=3087956;  
RA Todd J.A., Roberts A.N., Johnstone K., Piggot P.J., Winter G.,  
RA Ellar D.J.;  
RT "Reduced heat resistance of mutant spores after cloning and  
RT mutagenesis of the Bacillus subtilis gene encoding penicillin-binding  
RT protein 5.";  
RL J. Bacteriol. 167:257-264(1986).  
RN [5]  
RP SEQUENCE OF 414-443.  
RX MEDLINE=81117303; PubMed=6780559;  
RA Waxman D.J., Strominger J.L.;  
RT "Primary structure of the COOH-terminal membranous segment of a  
RT penicillin-sensitive enzyme purified from two Bacilli.";  
RL J. Biol. Chem. 256:2067-2077(1981).  
CC -!- FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE  
CC CELL WALL PRECURSORS.  
CC -!- CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O = 2 D-alanine.  
CC -!- PATHWAY: Peptidoglycan synthesis; final stages.  
CC -!- SUBCELLULAR LOCATION: Membrane-associated.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11; ALSO KNOWN AS THE  
CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; D26185; BAA05246.1; -;  
DR EMBL; Z99104; CAB11786.1; -;  
DR EMBL; M13766; AAA22375.1; -;  
DR PIR; S66040; S66040.  
DR MEROPS; S11.001; -;  
DR SubtilList; BG10074; dacA.  
DR InterPro; IPR001967; Ala/AlaCptasel.  
DR Pfam; PF00768; Peptidase\_S11; 1.  
DR PRINTS; PR00725; DADACBPTASE1.  
DR Hydrolase; Carboxypeptidase; Peptidoglycan synthesis; Cell wall;  
KW Membrane; Signal; Complete proteome.  
FT SIGNAL 1 31  
FT CHAIN 32 443 D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.  
FT ACT\_SITE 67 67 ACYLATED BY PENICILLIN.  
FT CONFLICT 100 100 E -> Q (IN REF. 3).  
FT CONFLICT 227 227 E -> Q (IN REF. 4).  
SQ SEQUENCE 443 AA; 48636 MW; DA6C5B0307D7C117 CRC64;  
Query Match 11.1%; Score 88.5; DB 1; Length 443;  
Best Local Similarity 24.8%; Pred. No. 11;  
Matches 31; Conservative 23; Mismatches 44; Indels 27; Gaps 4;  
QY 4 EVSELKHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQPDGWEISGEGKQDAG 63  
DB 318 EGDQVKGHK-TISVDKGEKEVGIVTKAFSLPVKNGE-EKNYKAK----- 361  
QY 64 YVINLSKDTFTFKPVFKKIEEKEENKPTFDVSKKQPNVHNSHRSKEDLQREH 123  
DB 362 --VTLNKDLNTPVKVKGKTKVGK-----LTAEYTGDEKDYGLNSDLGVDLVTKEN 410  
QY 124 SQKSD 128  
DB 411 VEKAN 415  
RESULT 10  
IGA2\_HAEIN STANDARD; PRT; 1702 AA.  
AC P45384;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).  
GN IGA.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HK715 / Serotype B;  
RX MEDLINE=92234949; PubMed=1373717;  
RA Poulsen K., Reinholdt J., Kilian M.;  
RT "A comparative genetic study of serologically distinct Haemophilus  
RT influenzae type 1 immunoglobulin A1 proteases.";  
RL J. Bacteriol. 174:2913-2921(1992).  
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
CC PRODUCING INTERACT FC AND FAB FRAGMENTS.  
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule  
CC substrates are known.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY

CC CC SIMILARITY).

CC CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.

CC CC -----

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC CC the European Bioinformatics Institute. There are no restrictions on its

CC CC use by non-profit institutions as long as its content is in no way

CC CC modified and this statement is not removed. Usage by and for commercial

CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC CC -----

CC CC EMBL; M87489; AA24966.1; -

CC CC PIR; A41859; A41859.

CC CC MEROPS; S06.001; -

CC CC InterPro; IPR006315; Autotransporter.

CC CC InterPro; IPR005546; Autotransporter.

CC CC InterPro; IPR000710; IGA\_S6.

CC CC InterPro; IPR004899; Pertactin.

CC CC Pfam; PF02395; IGA1; 1.

CC CC Pfam; PF03212; Pertactin; 1.

CC CC PRINTS; PR00521; IGASERTASE.

CC CC TIGRFAMs; TIGR01414; autotrans barl; 1.

CC CC Hydroxylase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.

CC CC SIGNAL 1 25

CC CC CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.

CC CC PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).

CC CC ACT\_SITE 288 288 PROBABLE.

CC CC DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-

CC CC K.

CC CC REPEAT 1109 1116 1.

CC CC REPEAT 1117 1124 2.

CC CC SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

CC CC

Query Match 11.0%; Score 87.5; DB 1; Length 1702;

Best Local Similarity 27.2%; Pred. No. 55;

Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 66 INLSKDTFIKVPFKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHS- 124

Db 1296 INTGSATAITAEKSDKQPQTETAASTEDASQHKANTVADNSVANNSESEPKRRRSI 1355

Qy 125 -QKSDSTKDVATVLDKNNISSTNNPK 154

Db 1356 SQPQTSAEETTAASDTETTDIADNSKRSKPNR 1387

RESULT 11

GYRA\_STAEP STANDARD; PRT; 893 AA.

AC P54112;

DT 01-OCT-1996 (Rel. 34, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE DNA gyrase subunit A (EC 5.99.1.3).

GN GYRA OR SE0005.

OS Staphylococcus epidermidis.

OC Bacteria; Firmicutes; Bacillales; Staphylococcus.

OX NCBI\_TaxID=1282;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 12228;

RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,

RA Chen Z., Wen Y.;

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

[2]

RP SEQUENCE OF 1-94 FROM N.A. AND MUTAGENESIS OF SER-84.

RX MEDLINE=92102204; PubMed=1662027;

RA Sreedharan S., Peterson L.R., Fisher L.M.;

RT "Ciprofloxacin resistance in coagulase-positive and -negative

RT staphylococci: role of mutations at serine 84 in the DNA gyrase A

RT protein of Staphylococcus aureus and Staphylococcus epidermidis.";

RL Antimicrob. Agents Chemother. 35:2151-2154(1991).

CC CC -!- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-

CC CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE

CC CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED

CC CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.

CC CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining

CC CC of double-stranded DNA.

CC CC -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA

CC CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE

CC CC ENZYME FORMS AN A2B2 TETRAMER.

CC CC -----

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC CC the European Bioinformatics Institute. There are no restrictions on its

CC CC use by non-profit institutions as long as its content is in no way

CC CC modified and this statement is not removed. Usage by and for commercial

CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC CC -----

CC CC EMBL; AE016744; AA003602.1; -

CC CC EMBL; S72603; AAB20672.1; -

CC CC PIR; A49832; A49832.

CC CC HSSP; P03097; 1A84.

CC CC InterPro; IPR002205; DNA\_topoisolv.

CC CC Pfam; PF03989; DNA\_gyraseA\_C; 6.

CC CC ProDom; PD000742; DNA\_topoisolv; 1.

CC CC SMART; SM00434; TOP4C; 1.

CC CC TIGRFAMs; TIGR01063; GYRA; 1.

CC CC Topoisomerase; Isomerase; DNA-binding; Antibiotic resistance;

CC CC Complete proteome. 123 123 DNA CLEAVAGE (BY SIMILARITY).

CC CC ACT\_SITE 84 84 S->F: RESISTANT TO CIPROFLOXACIN.

CC CC MUTAGEN 84 84

CC CC SEQUENCE 893 AA; 100113 MW; 2A6A7CD345A526CE CRC64;

CC CC

Query Match 10.9%; Score 87; DB 1; Length 893;

Best Local Similarity 25.3%; Pred. No. 30;

Matches 38; Conservative 27; Mismatches 65; Indels 20; Gaps 7;

Qy 14 TVTTONGKEMS-STIVSEEDFILPVKGEKGYQFDGWEISGFEKKDAGY-VINLSKD 71

Db 746 TITERNGNIVCITVTGEEDLMVTNAGVI---TLDVHDISQ-NGRAAQGVRLMKLGDG 801

Qy 72 TFIKPVFKKIEKKEEN-----KPTFDVSKKDNQVNHSQLNESHKEDLQREE 122

Db 802 QFVSTVAKVNEEDDNEENADEAQOSTTTETADVEVVD---DQTPGNAIHTEGDAEMES 857

Qy 123 -HSOKSDSTKDVATVLDKNNISSTNN 151

Db 858 VESPENDRIDRIDQDFMDRVNEDIESASDN 887

RESULT 12

MAPB\_MOUSE

ID MAPB\_MOUSE STANDARD; PRT; 2464 AA.

AC P14873;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))

DE [Contains: MAP1 light chain LC1].

GN MAP1B OR MTAP1B OR MTAP5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A. AND DOMAIN.

RC STRAIN=Swiss Webster; Tissue=Brain;

RX MEDLINE=90094539; PubMed=2480963;

RA Noble M., Lewis S.A., Cowan N.J.;

RT "The microtubule binding domain of microtubule-associated protein

RT MAP1B contains a repeated sequence motif unrelated to that of MAP2

RT and tau.";

RL J. Cell Biol. 109:3367-3376(1989).

CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS  
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
 CC STABILIZING MICROTUBULES.  
 CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
 CC WITH MAP1A AND MAP1B PROTEINS.  
 CC -!- DOMAIN: Has a highly basic region with many copies of the sequence  
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
 CC responsible for the binding of MAP1B to microtubules.  
 CC -!- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED  
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH  
 CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION  
 CC OF MAP1B.  
 CC -!- SIMILARITY: TO MAP1A.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X51396; CAA35761.1; -.  
 CC PIR; S07549; QRMSP1.  
 CC MGD; MGI:1306778; Mtap1b.  
 CC GO; GO:0016358; P:dendrite morphogenesis; IMP.  
 CC GO; GO:0001578; P:microtubule bundling; IMP.  
 CC InterPro: IPR000102; MAP1B\_neuraxin.  
 CC Pfam; PF00414; MAP1B\_neuraxin; 10.  
 CC PROSITE; PS00230; MAP1B\_NEURAXIN; 7.  
 CC Microtubules; Repeat; Phosphorylation.  
 CC CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.  
 CC REPEAT 1874 1890 MAP1B 1.  
 CC REPEAT 1891 1907 MAP1B 2.  
 CC REPEAT 1908 1924 MAP1B 3.  
 CC REPEAT 1925 1941 MAP1B 4.  
 CC REPEAT 1942 1958 MAP1B 5.  
 CC REPEAT 1959 1975 MAP1B 6.  
 CC REPEAT 1993 2009 MAP1B 7.  
 CC REPEAT 2010 2026 MAP1B 8.  
 CC REPEAT 2027 2043 MAP1B 9.  
 CC REPEAT 2044 2060 MAP1B 10.  
 CC DOMAIN 589 787  
 CC  
 CC SEQUENCE 2464 AA; 270408 MW; PBD3DD99CFDBDA87 CRC64;  
 CC  
 CC Query Match 10.9%; Score 87; DB 1; Length 2464;  
 CC Best Local Similarity 28.4%; Pred. No. 89;  
 CC Matches 29; Conservative 16; Mismatches 37; Indels 20; Gaps 4;  
 CC  
 CC QY 68 LSKDTFKVFK-KTEEKKEENKPTFDYKKKNDPQVNHQSQINSHRKEI-ORE----- 121  
 CC DB 632 VTNDKVVKKIETKLEKEE--KPKKEVVKEDKTPL---KXDEKPRKEEVKKEIKKEI 686  
 CC QY 122 -----EHSQKSDSTKDTVTATVLDKNNISKSTTNPNK 154  
 CC DB 687 KKEERKELKKEVKETPLDKAKKEVKKEEKEVKKEEKK 728  
 CC  
 CC ID RAT1 YEAST STANDARD; PRT; 1006 AA.  
 CC AC Q02782;  
 CC DT 01-OCT-1993 (Rel. 27, Created)  
 CC DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 CC DE Ribonucleic acid trafficking protein 1 (5'-3' exoribonuclease)  
 CC DE (EC 3.1.11.-) (P116).  
 CC GN RAT1 OR HKI1 OR TAP1 OR YOR048C.  
 CC OS Saccharomyces cerevisiae (Baker's yeast).

CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC NCBI\_TaxID=4932;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC RX MEDLINE=92331925; PubMed=1628825;  
 CC RA Amberg D.C., Goldstein A.L., Cole C.N.;  
 CC "Isolation and characterization of RAT1: an essential gene of  
 CC Saccharomyces cerevisiae required for the efficient nucleocytoplasmic  
 CC trafficking of mRNA.";  
 CC RT Genes Dev. 6:1173-1189 (1992).  
 CC [2]  
 CC SEQUENCE FROM N.A. AND EXORIBONUCLEASE ACTIVITY.  
 CC RX MEDLINE=93109318; PubMed=8417335;  
 CC RA Kenna M., Stevens A., McCammon M., Douglas M.G.;  
 CC "An essential yeast gene with homology to the exonuclease-encoding  
 CC XRN1/KEM1 gene also encodes a protein with exoribonuclease  
 CC activity.";  
 CC RT Mol. Cell. Biol. 13:341-350 (1993).  
 CC [3]  
 CC SEQUENCE FROM N.A.  
 CC RX MEDLINE=93268292; PubMed=8497260;  
 CC RA Aldrich T.L., di Segni G., McConaughy B.L., Keen N.J., Whelen S.,  
 CC Hall B.D.;  
 CC "Structure of the yeast TAP1 protein: dependence of transcription  
 CC activation on the DNA context of the target gene.";  
 CC RT Mol. Cell. Biol. 13:3434-3444 (1993).  
 CC [4]  
 CC SEQUENCE FROM N.A.  
 CC RX Landt O., Hiesel R., Unsel M.;  
 CC RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC [5]  
 CC SEQUENCE FROM N.A.  
 CC RA Bohn C., Bolotin-Fukuhara M., Daiguan-Fornier B., Dang D.V.,  
 CC Valens M.;  
 CC RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: May function in the processing and/or trafficking of  
 CC nuclear mRNA. May be involved in general transcription as well.  
 CC Possesses 5'->3' exoribonuclease activity degrading poly (A) to  
 CC mainly 5'-AMP.  
 CC -!- COFACTOR: Requires magnesium.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: Belongs to the 5'-3' exonuclease family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; S61567; AAB26818.1; -.  
 CC DB EMBL; M95626; AAA34960.1; -.  
 CC DB EMBL; L06011; AAA16950.1; -.  
 CC DB EMBL; Z11746; -; NOT ANNOTATED\_CDS.  
 CC DB EMBL; Z74956; CAA99240.1; -.  
 CC PIR; S20126; S20126.  
 CC SGD; S0005574; RAT1.  
 CC GO; GO:0005634; C:nucleus; IDA.  
 CC GO; GO:0004534; F:5'-3' exoribonuclease activity; IDA.  
 CC GO; GO:0006365; P:35S primary transcript processing; IMP.  
 CC GO; GO:0006396; P:RNA processing; IMP.  
 CC InterPro; IPR004859; Put 53exo.  
 CC Pfam; PF03159; XRN N; 1.  
 CC Nuclear protein; Hydrolase; Nuclease; Exonuclease; Repeat.  
 CC DOMAIN 525 528 POLY-GLU.  
 CC FT DOMAIN 793 797 POLY-ASN.  
 CC FT DOMAIN 955 999 CONTAINS 2 X SYRD, 2 X NNNY, AND 2 X  
 CC Y->H; IN ALLELE TAP1-1; ACTIVATES  
 CC TRANSCRIPTION OF THE PROMOTER-DEFECTIVE  
 CC YEAST SUP4 TRNA (TYR) ALLELE SUP4A53T61.  
 CC FT MUTAGEN 683 683  
 CC FT



Query Match 10.8%; Score 86; DB 1; Length 443;  
 Best Local Similarity 17.8%; Pred. No. 17;  
 Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;

QY 38 YKGELEKGYQDFGWEISGEGKADAGYVNLKDTFI-----74  
 DB 42 YEAVEDGVSFGSSIPGEGIEDSLIFKADPSTVAEIPWEGIGRVYIYKGBEPYQA 101  
 QY 75 --KPVFKKIEEKEEN-----KPTFDVSKKKD-----100  
 DB 102 DFRGILKRVLERLEKGLKALHIGPPEFYFKKNGTWELHPDSGGYFDLVGLDKAREIR 161  
 QY 101 -----NPNVNSHLSHREKEDLQREHSQKSD-----STKDVATATVLD 139  
 DB 162 REIALYMPYLGKLPVILHHEVGAKAHEIDFRYDEALRTADNIVSFKHVVKAAVE 215

## RESULT 16

MDNI\_HUMAN STANDARD; PRT; 5596 AA.  
 AC Q9NU22; O15019;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Midasin (MIDAS-containing protein).  
 GN MDNI OR KIAA0301.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX PubMed=12102729;  
 RA Garbarino J.E., Gibbons I.R.;  
 RT "Expression and genomic analysis of midasin, a novel and highly  
 conserved AAA protein distantly related to dynein.";  
 RL BMC Genomics 3:18-18(2002).  
 RN [2]  
 RP SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.  
 RC TISSUE=Testis;  
 RA Tracey A.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 3550-5596 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 The complete sequences of 100 new cDNA clones from brain which can  
 code for large proteins in vitro.";  
 RL DNA Res 4:141-150(1997).  
 CC -!- FUNCTION: May function as a nuclear chaperone and be involved in  
 the assembly/disassembly of macromolecular complexes in the  
 nucleus.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- SIMILARITY: Contains 1 WWFA domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF503925; AAM77722.1; -  
 DR EMBL; AL096678; CAB86660.1; -  
 DR EMBL; AL096678; CAB86661.1; -  
 DR EMBL; AB002299; BAA20761.1; -  
 DR Genew; HGNC:18302; MDNI.

DR GO; GO:0005634; C:nucleus; NAS  
 DR GO; GO:0016887; F:ATPase activity; NAS.  
 DR GO; GO:0003754; F:chaperone activity; NAS.  
 DR GO; GO:0006461; F:protein complex assembly; NAS.  
 DR InterPro; IPR003593; AAA ATPase.  
 DR InterPro; IPR002035; VWFA\_A.  
 DR Pfam; PF00004; AAA; 1.  
 DR SMART; SM00382; AAA; 7.  
 DR SMART; SM00327; VWFA; 1.  
 DR PROSITE; PS0234; VWFA; 1.  
 KW Chaperone; ATP-binding; Repeat; Nuclear protein.  
 FT NP\_BIND 329 336 ATP (POTENTIAL).  
 FT NP\_BIND 677 684 ATP (POTENTIAL).  
 FT NP\_BIND 1084 1091 ATP (POTENTIAL).  
 FT NP\_BIND 1390 1397 ATP (POTENTIAL).  
 FT NP\_BIND 1753 1760 ATP (POTENTIAL).  
 FT NP\_BIND 2066 2073 ATP (POTENTIAL).  
 FT DOMAIN 3566 3573 POLY-GLU.  
 FT DOMAIN 4784 4791 POLY-GLU.  
 FT DOMAIN 5008 5013 POLY-GLU.  
 FT DOMAIN 5182 5187 POLY-GLU.  
 FT DOMAIN 5384 5583 VWFA.  
 FT CONFLICT 2287 2312 RLEFLSMDPVHGDISRMRNRLGIYI -> S (IN REF.  
 FT 2).  
 SQ SEQUENCE 5596 AA; 632802 MW; 586C626161F96D4 CRC64;  
 Query Match 10.8%; Score 86; DB 1; Length 5596;  
 Best Local Similarity 23.4%; Pred. No. 2.5e+02;  
 Matches 30; Conservative 27; Mismatches 63; Indels 8; Gaps 2;

QY 8 LKPHRVTVTIQNGKMSSTIVSEEDFILPVVYKGELEK-----YQFDGWEISGEGKDD 61  
 DB 4629 LATHRSTAKLLSLVAQVFTLAQKGFCLPKFEMDSAGEGATEFHDYEGGIGEGEKMD 4689  
 QY 62 AGYVINLSKDTFIRPKKIEEKEENKPTFDVSKKDNPNVNSHLSHREKEDLQRE 121  
 DB 4689 VS--DQIGNEQVEDTFQGEKQKEDPDSKDKNGEDNAIEMSEDFQKWHGDGELEGE 4745  
 QY 122 EHSQKSDS 129  
 DB 4747 EDDEKSDS 4754

RESULT 17  
 PBPA\_BACSU STANDARD; PRT; 914 AA.  
 AC P39793;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Penicillin-binding protein 1A/1B (PBP1) [includes: Penicillin-  
 insensitive transglycosylase (EC 2.4.2.-) (peptidoglycan TGase);  
 Penicillin-sensitive transpeptidase (EC 3.4.-.-) (DD-transpeptidase)].  
 GN PONA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 499-515.  
 RC STRAIN=168;  
 RX MEDLINE=95113769; PubMed=7814321;  
 RA Popham D.L., Setlow P.;  
 RT "Cloning, nucleotide sequence, and mutagenesis of the Bacillus  
 subtilis pona operon, which codes for penicillin-binding protein  
 (PBP) 1 and a PBP-related factor.";  
 RL J. Bacteriol. 177:326-335 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / Marburg;  
 RX MEDLINE=96349105; PubMed=8760912;  
 RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,  
 RA Serron P.;  
 RT "Sequence analysis of the Bacillus subtilis chromosome region between





OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
RC STRAIN=S288c; PubMed=8657111;  
RX MEDLINE=96239492; PubMed=8657111;  
RA Chen G.-C., Zheng L., Chan C.S.M.;  
RT "The LIM domain-containing Dbp1 GTPase-activating protein is required  
RT for normal cellular morphogenesis in *Saccharomyces cerevisiae*.";  
RL Mol. Cell. Biol. 16:1376-1390(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / FY1679;  
RX MEDLINE=97060020; PubMed=8904341;  
RA Wiemann S., Rechmann S., Benes V., Voss H., Schwager C., Vlcek C.,  
RT Stegemann J., Zimmermann J., Erfle H., Paces V., Ansoerge W.;  
RT "Sequencing and analysis of 51 kb on the right arm of Chromosome XV  
RT from *Saccharomyces cerevisiae* reveals 30 open reading frames.";  
RL Yeast 12:281-288(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97344368; PubMed=9200815;  
RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,  
RT Schwager C., Paces V., Sander C., Ansoerge W.;  
RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV.";  
RL Yeast 13:655-672(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c;  
RX MEDLINE=96101594; PubMed=7498791;  
RA Stevenson B.J., Ferguson B., de Virgilio C., Bi E., Pringle J.R.,  
RT Ammerer G., Sprague G.F. Jr.;  
RT "Mutation of RGA1, which encodes a putative GTPase-activating protein  
RT for the polarity-establishment protein Cdc42p, activates the  
RT pheromone-response pathway in the yeast *Saccharomyces cerevisiae*.";  
RL Genes Dev. 9:2949-2963(1995).  
RN [5]  
RP SEQUENCE OF 570-639 FROM N.A.  
RC STRAIN=SNY243;  
RX MEDLINE=93087574; PubMed=1454852;  
RA Ramer S.W., Elledge S.J., Davis R.W.;  
RT "Dominant genetics using a yeast genomic library under the control of  
RT a strong inducible promoter.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:11589-11593(1992).  
CC -1- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR CDC42 AND/OR RHO1.  
CC NEGATIVE REGULATOR OF THE PHEROMONE-RESPONSE PATHWAY THROUGH THE  
CC STE20 PROTEIN KINASE. ACTS AT A STEP BETWEEN THE G-PROTEIN AND THE  
CC MAP KINASE MODULE. DOMINANT SUPPRESSOR OF BUD EMERGENCE DEFECT  
CC CAUSED BY DELETION OF IPL2/BEM2. INVOLVED IN THE CONTROL OF  
CC POLARIZED CELL GROWTH AND PROPER BUD SITE SELECTION.  
CC -1- SIMILARITY: Contains 2 LIM zinc-binding domains.  
CC -1- SIMILARITY: Contains 1 Rho-GAP domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U07421; AAA16875.1; -  
CC EMBL; X90518; CAA62108.1; -  
CC EMBL; X94335; CAA64046.1; -  
CC EMBL; Z75035; CAA9326.1; -  
CC EMBL; X90950; CAA62445.1; -  
CC EMBL; L02617; AAA35153.1; -  
CC PIR; S48535; S48535.  
CC SGD; S0005653; RGA1.  
CC GO; GO:0005100; F:Rho GTPase activator activity; IPI.  
CC GO; GO:0004871; F:signal transducer activity; IPI.  
CC GO; GO:0007015; P:actin filament organization; IMP.

DR GO; GO:0007118; P:apical bud growth; IPI.  
DR GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. . .; IPI.  
DR GO; GO:0007125; P:invasive growth; IPI.  
DR GO; GO:0007119; P:isotropic bud growth; IPI.  
DR GO; GO:0007124; P:pseudohyphal growth; IPI.  
DR GO; GO:0000750; P:signal transduction during conjugation with. . .; IGI.  
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IPI.  
DR InterPro; IPR001781; LIM.  
DR InterPro; IPR000198; RhoGAP.  
DR Pfam; PF00412; LIM; 1.  
DR Pfam; PF00620; RhoGAP; 1.  
DR ProDom; PD000094; LIM; 2.  
DR SMART; SM00132; LIM; 2.  
DR SMART; SM00324; RhoGAP; 1.  
DR PROSITE; PS00478; LIM DOMAIN\_1; 1.  
DR PROSITE; PS0023; LIM DOMAIN\_2; 2.  
DR PROSITE; PS0238; RhoGAP; 1.  
KW GTPase activation; Repeat; LIM domain; Metal-binding; Zinc;  
KW Pheromone response.  
FT DOMAIN 13 66 LIM 1.  
FT DOMAIN 70 122 LIM 2.  
FT DOMAIN 791 1006 RHO-GAP.  
FT VARIANT 866 866 V -> A.  
FT VARIANT 898 898 K -> R.  
FT VARIANT 926 926 S -> G.  
FT MUTAGEN 40 40 C->S: BIPOLAR BUDDING.  
FT MUTAGEN 37 37 C->S: BIPOLAR BUDDING.  
FT MUTAGEN 98 98 C->S: BIPOLAR BUDDING.  
FT MUTAGEN 101 101 C->S: BIPOLAR BUDDING.  
FT CONFLICT 457 457 D -> E (IN REF. 4).  
FT CONFLICT 507 507 T -> P (IN REF. 4).  
SQ SEQUENCE 1007 AA; 112831 MW; C805411B57553791 CRC64;  
Query Match 10.7%; Score 85.5; DB 1; Length 1007;  
Best Local Similarity 21.0%; Pred. No. 44;  
Matches 34; Conservative 34; Mismatches 69; Indels 25; Gaps 3;  
QY 1 DTGEVSELKPHRVTVTQNGKMSSTI-----VSEDFILPVYGELEKGYQFGWEIS 54  
DB 349 NTGEISQMDPSLSRKVLNNIVEETNALQRPVVVVKEDRSVPDLAGVQOQAEKYSYNN 408  
QY 55 GPEGKDGAGYVNLKSDTFIKPVPFKIEKKEENKPTFDVSKKDNPOVNHSQLNESH 114  
DB 409 SGKGRKISRSLRRSRKDLMI-----NLKSRATCKQDSNVKLSPAKVTSR 454  
QY 115 KEDLOREHSQKSDSTKDTATVLD-----KNNISKSTTNN 151  
DB 455 SODLRDNDSDHTGLDTPNSNSTSLDLVNNQKSLNVKFTDN 496  
RESULT 19  
ID IF2\_STAEP STANDARD; PRT; 720 AA.  
AC Q8CST4.  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Translation initiation factor IF-2.  
GN INFB OR SE0945.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
RA Chen Z., Wen Y.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBSJ databases.  
CC -1- FUNCTION: One of the essential components for the initiation of  
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous  
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.  
CC Also involved in the hydrolysis of GTP during the formation of the  
CC 70S ribosomal complex (By similarity).

```

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016747; AA004542.1; -
DR HAMAP; MF_00100; -; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR Pfam; PF04760; IF2_N_2; 1.
DR ProDom; PD186100; IF2; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 225 373 G-DOMAIN.
FT NP_BIND 231 238 GTP (BY SIMILARITY).
FT NP_BIND 277 281 GTP (BY SIMILARITY).
FT NP_BIND 331 334 GTP (BY SIMILARITY).
SQ SEQUENCE 720 AA; 79343 MW; 07F55A6A59CF970C CRC64;

Query Match 10.6%; Score 85; DB 1; Length 720;
Best Local Similarity 28.0%; Pred. NO. 34;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

Qy 81 IEEKEENKPTEDVSKKD---NPQVNSQLNSHREKDLQREHSQKSDSTK---D 132
Db 38 LEEEOIKALDKKFKASQASQAKNTQNTQNNHOKSNKQNSDKRQKQSKNKPTKKEQN 97

Qy 133 VTATVLDKNKNSKSTTNNPNK 154
Db 98 NKGQKQNNKNTKNQKNNK 119

RESULT 20
ID UN89 CAEEL STANDARD; PRT; 6632 AA.
AC 001761; Q17362;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OX Rhabditidae; Pelodierinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]_TaxID=6239;
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT assembly, encodes a giant modular protein composed of Ig and signal
RT transduction domains";
RL J. Cell Biol. 132:835-848 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofilament

```

```

CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U33058; AAB00542.1; -
DR EMBL; AF003131; AAB54132.2; -
DR PDB; 1FHO; 20-DEC-00.
DR WormPep; C09D1.1; CE30426.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR007110; IG-Like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001849; PH.
DR InterPro; IPR007850; RCSD.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 47.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF05177; RCSD; 5.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IG2; 23.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS50835; IG_LIKE; 49.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.

```

```
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4201 4297 IG-LIKE C2-TYPE 28.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 29.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 30.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 31.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 32.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 33.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 34.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 35.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 36.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 37.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 38.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 39.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 40.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 41.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 42.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 43.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 44.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 45.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 46.
FT DOMAIN 6275 6368 IG-LIKE C2-TYPE 47.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3085 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFLICT 2137 2137 A -> P (IN REF. 1).
FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).
FT CONFLICT 2258 2258 A -> P (IN REF. 1).
FT CONFLICT 2284 2284 E -> G (IN REF. 1).
FT CONFLICT 2297 2297 M -> I (IN REF. 1).
FT CONFLICT 3531 3531 A -> G (IN REF. 1).
FT CONFLICT 3884 3888 DAGGY -> RRRRI (IN REF. 1).
FT CONFLICT 3929 3929 A -> V (IN REF. 1).
FT CONFLICT 5134 5134 A -> P (IN REF. 1).
FT CONFLICT 5145 5145 T -> S (IN REF. 1).
FT CONFLICT 5185 5185 G -> A (IN REF. 1).
FT CONFLICT 5199 5199 K -> N (IN REF. 1).
FT CONFLICT 5202 5202 L -> F (IN REF. 1).
FT CONFLICT 5213 5213 F -> L (IN REF. 1).
FT CONFLICT 6178 6178 A -> G (IN REF. 1).
FT CONFLICT 6268 6268 K -> E (IN REF. 1).
SQ SEQUENCE 6632 AA; 731665 MW; 262D3EDD62960E89 CRC64;

Query Match 10.6%; Score 85; DB 1; Length 6632;
Best Local Similarity 22.1%; Pred. No. 3.6e+02;
Matches 34; Conservative 32; Mismatches 68; Indels 20; Gaps 7;

QY 7 ELKPHRVTVTQNGKMSSTIVSEEDF--ILPVYKGELEKGYQFDGWE--ISGPEGKGD 61
DB 1143 DLRP--VSLTVKSGE---AVFSAHAFGLPLTYEWSVNGRKVRDQGEARVTRDESTVD 1197
```



```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D49676; BAA08532.1; -
CC EMBL; U51224; AAA98669.1; -
CC MIM; 601079; -
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0003723; F:RNA binding activity; NAS.
CC InterPro; IPR000504; RNA_rec_mot.
CC InterPro; IPR000571; Znf_CCHC.
CC Pfam; PF00076; rrm; 1.
CC Pfam; PF00642; zf-CCHC; 2.
CC SMART; SM00360; RRM; 1.
CC SMART; SM00356; ZNF_C3H1; 2.
CC PROSITE; PS50102; RRM; 1.
CC PROSITE; PS00030; RRM_RNP_1; FALSE NEG.
CC NUCLEAR protein; RNA-binding; Ribonucleoprotein; Zinc-finger; Repeat.
CC DOMAIN 203 309 96F326694BD4B7C0 CRC64;
CC SEQUENCE 479 AA; 57643 MW; 96F326694BD4B7C0 CRC64;

Query Match 10.5%; Score 83.5; DB 1; Length 479;
Best Local Similarity 21.7%; Pred. No. 28;
Matches 26; Conservative 25; Mismatches 40; Indels 29; Gaps 4;

QY 60 KDAGYVINTSKQFIRKVFVKIIEKK-----BEENKPFVSKKKDNQ 103
DB 44 RDSGLSQEEDFDFFE--QGLEEKLERERERLHEWLLRQKQOEERIKKEBEA 101
QY 104 VN-----HSQINSHREKEDLQREHSHQSDSTQVATATLD--KNNISKSTNNP 152
DB 102 KKLLEGERLTKKQWQKQKQKKEKEBEERQKQKQKKEKEANQKMLDQANDLENSTWQNP 161

RESULT 24
ARs2 DROME STANDARD; PRT; 943 AA.
ID ARs2 DROME STANDARD; PRT; 943 AA.
AC 09V9K7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arsenite-resistance protein 2 homolog.
GN CG7843.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wotman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Beeson K.Y., Basu P.V., Berman J.P., Bayraktaroglu L., Beasley E.M.,
RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```

```

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mikhlov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RA Science 287:2185-2195(2000).
RA [2]
RA REVISIONS, AND ALTERNATIVE SPLICING.
RA STRAIN=Berkeley;
RA MEDLINE=22426069; PubMed=12537572;
RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Procinik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RA "Annotation of the Drosophila melanogaster euchromatic genome: a
RA systematic review.";
RA Genome Biol. 3:RESEARCH0083.22(2002).
RA CC -1- ALTERNATIVE PRODUCTS:
RA Event=Alternative splicing; Named isoforms=2;
RA Name=Long;
RA IsoId=Q9V9K7-1; Sequence=Displayed;
RA Name=Short;
RA IsoId=Q9V9K7-2; Sequence=VSP 000327;
RA CC -1- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE003784; AAM68343.1; -
CC EMBL; AE003784; AAM68345.1; -
CC FlyBase; FBgn0033062; CG7843.
CC Pfam; PF04959; ARS2; 1.
CC Hypothetical protein; Alternative splicing.
CC VARSPLIC 47 50 Missing (in isoform short).
CC FT /ftid=VSP 000327.
CC FT SEQUENCE 943 AA; 107221 MW; 0CIAF09E02E8AB0B CRC64;

Query Match 10.5%; Score 83.5; DB 1; Length 943;
Best Local Similarity 25.0%; Pred. No. 58;
Matches 36; Conservative 23; Mismatches 68; Indels 17; Gaps 4;

QY 4 EYSELKP-----HRTVTIIONGKEMSSRT-----VSEEDLIPYKGELEKGYQFDGW 51
DB 272 KYLDEKPKDPVYERAEQMSVKEKTTINPKKEMSSADYVSTQRKVRPNVNSGWN 331
QY 52 EISGEFGKDGAVINTSKQFIRKVFVKIIEKK-----KKEENKPFVSKKKDNQ 108
DB 332 DDDDAENSAKPKKELMADSDSDSKPKDQINKKTKRKRRNSDDSSSSSSSSSDEER 391

```

OY 109 LINESHREDLOREHSHQSDSTKD 132  
DB 392 LKEKYDVEDGJRAE--QKTEAKD 413

RESULT 25

IF31\_HUMAN STANDARD; PRT; 258 AA.  
AC 075822; Q9BUD2; Q9H8Q2;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Eukaryotic translation initiation factor 3 subunit 1 (eIF-3 alpha)  
DE (eIF3 p35) (eIF3j).  
GN EIF3J.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=99041954; PubMed=9822659;  
RA Block K.L., Vornlocher H.-P., Hersey J.W.B.;  
RT "Characterization of cDNAs encoding the p44 and p35 subunits of human  
RL J. Biol. Chem. 273:31901-31908(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC Tissue=Ovarian carcinoma;  
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,  
RA Wagaetsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y.,  
RA Niimura K., Iwayanagi T.;  
RT "NEBO human cDNA sequencing project";  
RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC Tissue=uterus;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueda T.B., Teshiguchi S., Cantinci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Nadeau A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Scherch A., Schin J.E., Jones J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RL human and mouse cDNA sequences";  
CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF  
CC METHIONYL-TRNAI AND MRNA.  
CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 12 DIFFERENT SUBUNITS (By  
CC similarity).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
CC EMBL; U97670; AAC78729.1; -.  
CC EMBL; AK023388; BAB14555.1; -.  
CC EMBL; BC002719; AAH02719.1; -.  
CC Genew; HGNC:3270; EIF3S1.  
CC GK; 075822; -.  
CC MIM; 603910; -.  
CC DR GO; GO:0005852; C:eukaryotic translation initiation factor 3. . . . TMS.  
CC DR GO; GO:0003743; F:translation initiation factor activity; TMS.  
CC DR GO; GO:0006446; P:regulation of translational initiation; TMS.  
CC KM Initiation factor; Protein biosynthesis.  
CC FT DOMAIN 2 29 32 POLY-ALA.  
CC FT DOMAIN 53 57 POLY-GLY.  
CC FT DOMAIN 218 224 POLY-ASP.  
CC FT CONFLICT 40 40 E -> G (IN REF. 1).  
CC FT CONFLICT 141 141 A -> T (IN REF. 2).  
CC SQ SEQUENCE 258 AA; 29062 MW; 8362423542445AA CRC64;  
OY Query Match 10.4%; Score 83; DB 1; Length 258;  
OY Best Local Similarity 25.4%; Freq. No. 16;  
OY Matches 30; Conservative 25; Mismatches 51; Indels 12; Gaps 4;  
DB 38 YKGELEKGYDFGWEISGFEGKXGAGY--VINLSKDTFIRPKKIEKKEENKPTPDV 95  
DB 39 WEGEDEEDVDKNDMDDEKKEAEVPEVKISK--KIAEKIKERQKQKROBEI 95  
OY 96 SKKDNPQ-----VNHSQINESHREDLOREHSHQSDSTKYDTATV--LDKNITSK 146  
OY 96 KKRLEPEPEPVLPPEBQLADKLRLKQLQESDLELAKETGVNNAVYGIDAMPSSR 153  
DB 96 KKRLEPEPEPVLPPEBQLADKLRLKQLQESDLELAKETGVNNAVYGIDAMPSSR 153

RESULT 26

T0NB\_HAEIN STANDARD; PRT; 270 AA.  
ID T0NB\_HAEIN  
AC P42872;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE T0NB protein.  
GN T0NB OR H10251.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NTM1 TN106;  
RX MEDLINE=94245357; PubMed=8188372;  
RA Jarosik G.P., Sanders J.D., Cope L.D., Muller-Eberhard U.,  
RA Hansen E.J.;  
RT "A functional t0nb gene is required for both utilization of heme and  
RL virulence expression by Haemophilus influenzae type b.";  
RN Infect. Immun. 62:2470-2477(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty R.A., Merrick J.M.,  
RA Kennedy K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weiman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RL Rd.";  
CC Science 269:496-512(1995).  
CC -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT  
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO





SCIN. Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_taxid:9606;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC TISSUE=Skin;  
 CC MEDLINE=22386257; PubMed=12477932;  
 CC Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 CC Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 CC Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 CC Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 CC Ditchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
 CC Stapleton W., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 CC Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 CC Rata S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 CC Roark S.A., McGown P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 CC Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 CC Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 CC Fahey J., Helton E., Kettelman M., Maan A.C., Shvachenko Y., Bouffard G.G.,  
 CC Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 CC Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 CC Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 CC Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 CC Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;  
 CC "Generation and initial analysis of more than 15,000 full-length  
 CC human and mouse cDNA sequences.";  
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 CC [2]  
 CC SEQUENCE OF 1-527 FROM N.A.  
 CC RA Kalicki J., Smith-Craig R.;  
 CC Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 CC [3]  
 CC SEQUENCE OF 248-715 FROM N.A.  
 CC TISSUE=Placenta;  
 CC Isegal T., Ora T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 CC Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 CC Tanase T., Nomura Y., Togaya S., Komai F., Hara R., Takeuchi K.,  
 CC Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 CC Akamatsu A., Nakamura Y., Nagahata K., Masuno Y., Oshima A.;  
 CC "NDO human cDNA sequencing project";  
 CC Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Ca(2+)-dependent actin filament-severing protein that is  
 CC presumed to have a regulatory function in exocytosis by affecting  
 CC the organization of the microfilament network underneath the  
 CC plasma membrane. In vitro, also has barbed end capping and  
 CC nucleating activities in the presence of Ca(2+).  
 CC -1- SIMILARITY: BELONGS TO THE VILIN/GELSOLIN FAMILY.  
 CC -1- SIMILARITY: Contains 6 gelsolin-like repeats.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, BC021090, AAH21090.1; -  
 CC EMBL, AC005281, AAD15423.1; -  
 CC EMBL, AK027778, BAB55361.1; -  
 CC HSSP, P02640, 2VIL.  
 CC InterPro: IPR001974; Gelsolin.  
 CC Pfam, PF00626, Gelsolin; 6.  
 CC PRINTS, PR00597, GELSOLIN.  
 CC Cysoskeleton; Actin-binding; Repeat; Calcium; Capping protein.  
 CC SMART, SMO0262; GEL; 6.  
 CC DOMAIN 1 363  
 CC REPEAT 364 715  
 CC REPEAT 27 76  
 CC REPEAT 148 188  
 CC REPEAT 265 307  
 CC GELSOLIN-LIKE 1.  
 CC GELSOLIN-LIKE 2.  
 CC GELSOLIN-LIKE 3.

	FT	REPEAT	398	451	GELSOLIN-LIKE 4.
	FT	REPEAT	523	564	GELSOLIN-LIKE 5.
	FT	REPEAT	626	668	GELSOLIN-LIKE 6.
	FT	SITE	112	119	POLYPHOSPHOINOSTITIDE BINDING (BY SIMILARITY).
	FT	SITE	138	146	POLYPHOSPHOINOSTITIDE BINDING (BY SIMILARITY).
	FT	CONFLICT	61	61	R -> H (IN REF. 2).
	FT	SEQUENCE	715 AA;	45FBE42CBCEBDD80 CRC64;	
		Query Match	10.4%;	Score 83;	DB 1; Length 715;
		Best Local Similarity	23.1%;	Pred. No. 47;	
		Matches	34;	Conservative	30; Mismatches 47; Indels 36; Gaps 5;
OY		12 RVTYTIONGKMSSTIYSSEDFLLPYVKGELEKYQFDGMISGFEEK-----KDAGYV	65		
Dd		262 RVTVAAEENPEPSMAMLIISECFILLD--HGAKQLFWKKGDANQEERKAAMKTAEFLQQ	319		
OY		66 INLSKDPFIK-----PVFKK-----IEEKKEENKPFPDVSKKK	99		
Dd		320 MNYSKNIQIVLPFGCGTFPIFKQPKWRDKDSGGSGKVYTRKVAQIKQIPPDASKLH	379		
OY		100 DNPQV--NHSQMESHRRKLQREEHS	124		
Dd		380 SSPQMAAQHMNVDDSGKVEIMRWENN	406		
		RESULT 29			
		DNL_I_CAEEL STANDARD; PRT; 847 AA.			
ID		DNL_I_CAEEL			
AC		027474:			
DT		01-NOV-1997 (Rel. 35, Created)			
DT		01-NOV-1997 (Rel. 35, Last sequence update)			
DT		28-FEB-2003 (Rel. 41, last annotation update)			
DE		DNA ligase [EC 6.5.1.1] (Polydeoxyribonucleotide synthase [ATP]).			
GN		LIG-1 OR C29A12.3.			
OS		Caenorhabditis elegans.			
OC		Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilida; Rhabditoidea;			
CC		Rhabdillidae; Peloderinae; Caenorhabditis.			
OX		NCB1_TaxID=6239;			
RN		[1]			
RP		SEQUENCE FROM N.A.			
RC		STRAIN-Bristol NZ;			
RA		Wilkinson J.;			
CC	-	Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.			
CC	-	FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA			
CC	-	RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRAWDED DNA.			
CC	-	CATALYTIC ACTIVITY: ATP + {deoxyribonucleotide} (N) +			
CC		{deoxyribonucleotide} (N+M).			
CC		-- -- -- -- --			
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC		between the Swiss Institute of Bioinformatics and the EMBL outpostation -			
CC		the European Bioinformatics institute. There are no restrictions on its			
CC		use by non-profit institutions as long as its content is in no way			
CC		modified and this statement is not removed. Usage by and for commercial			
CC		entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC		or send an email to license@lsb-sib.ch).			
CC		-- -- -- -- --			
DR		EMBL; Z73970; CA98242.1; -			
DR		PIR; T19544; T19544.			
DR		Morppep; C29A12.3; CE05328.			
DR		InterPro; IPR000977; DNA_ligase.			
DR		pfam; PF01068; DNA_ligase_1.			
DR		pfam; PF04679; DNA_ligase_A_C_1.			
DR		pfam; PF04675; DNA_ligase_A_N_1.			
DR		TIGRFAMS; TIGR00574; dnl1; 1.			
DR		PROSITE; PS00697; DNA_LIGASE A1; 1.			
DR		PROSITE; PS00333; DNA_LIGASE A2; 1.			
KW		PROSITE; PSS0160; DNA_LIGASE A3; 1.			
KW		DNA repeat; DNA replication; DNA recombination; Cell division; Ligase;			
FM		ATP-binding.			
FT	BINDING	363	363	AMP (BY SIMILARITY).	

SEQUENCE 847 AA; 94501 MW; A32B7D729F91002F CRC64;  
Query Match 10.4%; Score 83; DB 1; Length 847;  
Best Local Similarity 26.2%; Pred. No. 56;  
Matches 44; Conservative 26; Mismatches 68; Indels 30; Gaps 8;  
QY 1 DPGVESELKPHRVTVTIQNGKEMSTIVSEDFILPYKGE-----LEKGYQFGWEI 53  
DB 656 DDCGKISLRRPR-LIRDPKNSDDATSSQ--VLEMYKQGFAPNOKIEKADAVD--ED 710  
QY 54 SGFEGKDAAGYVNL-----SKDTFIKPVFKIEEKEENKPTFDVSKK--DNPOV 104  
DB 711 DEFEKEDEBEELMTNVSQSKENPVK-----EIKKETPK---SVSPKFEKKPPV 761  
QY 105 NNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNISSEKSTNNP 152  
DB 762 KSPVVKSPVKSPIKKEAKKKGPVASIFSSSTKKNEDKVKESDPS 809  
RESULT 30  
GLNA\_PYRAB STANDARD; PRT; 439 AA.  
ID GLNA\_PYRAB  
AC Q9UY99;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
GN GLNA OR PYRAB16090 OR PAB1292.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GS5 / Orsay.  
RX PubMed=12622808;  
RA Cohen G.N., Barbe V., Flament D., Galperin M., Hellig R., Lecompte O.,  
Roch O., Prieur D., Querellon J., Ripp R., Thierry J.-C.,  
Van der Oost J., Weissendach J., Zivanovic Y., Forterre P.;  
RT "An integrated analysis of the genome of the hyperthermophilic  
archaeon Pyrococcus abyssi.";  
RL Mol. Microbiol. 47:1495-1512(2003).  
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
L-glutamine.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
EMBL; AJ248288; CAB50513.1; -.  
DR PIR; C75009; C75009.  
DR HSSP; P06201; 1LGR.  
DR InterPro; IPR001691; GLN synth.  
DR InterPro; IPR004809; GLNA.  
DR InterPro; IPR001637; GLNA adenyltn.  
DR Pfam; PF00120; gln-synt; 1.  
DR Pfam; PF03951; gln-synt N; 1.  
DR ProDom; PD001057; gln-synt C; 1.  
DR TIGRPFAM; TIGR00653; GLNA; 1.  
DR PROSITE; PS00180; GLNA\_1; 1.  
DR PROSITE; PS00181; GLNA\_ATP; 1.  
KW Ligase; Complete proteome.  
FT BINDING 358  
FT AMP (UNDER CONDITIONS OF ABUNDANT  
FT GLUTAMINE) (BY SIMILARITY).  
SQ SEQUENCE 439 AA; 49742 MW; 62CCFD3970A98AF0 CRC64;  
Query Match 10.3%; Score 82.5; DB 1; Length 439;

Best Local Similarity 17.8%; Pred. No. 30;  
Matches 28; Conservative 24; Mismatches 36; Indels 69; Gaps 4;  
QY 42 LEKGYQFGDWEISGREGKADAGVNLISDTFI-----KP----- 76  
DB 42 IEDGISFDSSVPGQIEDSDLIFKADPDITVEVPMDNARVGYIKDGRPGADPRG 101  
QY 77 VFKEIEEKEEN-----KPTFDVSKKDN----- 101  
DB 102 VLKRVIEKLAENGIRAYIGPEPEFYLFKNGSWELEIPDVGYPDILITDKAKDIKREIA 161  
QY 102 -----POVNSQLNESHKEDLQREHSQKSDS 129  
DB 162 EYMPSEGLVPEVILHHVGKAQHEIDPRYDEALKTADN 198  
RESULT 31  
HS70\_PYRSA STANDARD; PRT; 649 AA.  
ID HS70\_PYRSA  
AC P37859;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Heat shock 70 kDa protein.  
GN HSP70.  
OS Pyrenomonas salina.  
OC Nucleomorph.  
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Pyrenomonas.  
OC NCBI\_TaxID=3034;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94268506; PubMed=8208251;  
RA Hofmann C.J.B., Rensing S.A., Hauber M.M., Martin W.F., Mueller S.B.,  
RA Couch J., McFadden G.I., Igloi G.L., Mäler U.-G.,  
RT "The smallest known eukaryotic genomes encode a protein gene: towards  
an understanding of nucleomorph functions.";  
RL Mol. Genet. 243:600-604(1994).  
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
EMBL; X72621; CAA51197.1; -.  
DR PIR; S42488; S42488.  
DR HSSP; P08109; 1CKR.  
DR InterPro; IPR001023; Hsp70.  
DR Pfam; PF00012; HSP70; 1.  
DR PRINTS; PR00301; HEATSHOCK70.  
DR ProDom; PD000089; HSP70; 1.  
DR PROSITE; PS00297; HSP70\_1; 1.  
DR PROSITE; PS00329; HSP70\_2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
KW ATP-binding; Heat shock; Nucleomorph.  
SQ SEQUENCE 649 AA; 72079 MW; B627B08F9C9164 CRC64;  
Query Match 10.3%; Score 82.5; DB 1; Length 649;  
Best Local Similarity 23.8%; Pred. No. 46;  
Matches 34; Conservative 25; Mismatches 43; Indels 41; Gaps 7;  
QY 9 KPHRVTVTIQNG-----KEMSTIVSEDFILPYKGELEKGYQFGDWEISGFEKDA--- 62  
DB 503 KSNKITITNDKRLSLSEIEHVRVEAE-----KYKTEDEK-----LKKLEAKNS 547  
QY 63 --GYVINS---KDTFIKPVFKIEEKEENKPTFDVSKKDNPOVNSQLNESHKED 117  
DB 548 LENVANININVTVD-----EKLKEKIQEEDKKSIIEKYKE-----VLEFIETNBD 592  
QY 118 LQREHSQKSDSTKDVATVLDK 140

DB 593 LEKEEYERKEKEKRNFPANPIISK 615

RESULT 32  
MAPB\_RAT STANDARD; PRT; 2459 AA.  
AC P15205: Q62958; Q9ER21; Q9QW92;  
01-APR-1990 (Rel. 14, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Microtubule-associated protein 1B (MAP1B) [Contains: MAP1  
light chain LC1].  
GN MAP1B.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE OF 1-142 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Testis;  
RX MEDLINE=96257242; PubMed=8666295;  
RA Liu D., Fischer I.;  
RT "Isolation and sequencing of the 5' end of the rat microtubule-  
associated protein (MAP1B)-encoding cDNA.";  
RL Gene 172:307-308(1996).  
RN [2]  
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.  
RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Gli1 tumor;  
RX MEDLINE=92347374; PubMed=1639092;  
RA Zanner W., Kratz J., Staunton J., Feick P., Wiche G.;  
RT "Identification of two distinct microtubule binding domains on  
recombinant rat MAP1B.";  
RL Eur. J. Cell Biol. 57:66-74(1992).  
RN [3]  
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.  
RC TISSUE=Spinal cord;  
RX MEDLINE=90059671; PubMed=2555150;  
RA Lientz A., Gremmling G., Hermans-Borgmeyer I., Kirsch J.,  
RA Lientz U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;  
RT "Neuraxin, a novel putative structural protein of the rat central  
nervous system that is immunologically related to microtubule-  
associated protein 5.";  
RL EMBO J. 8:2879-2888(1989).  
RN [4]  
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.  
RX MEDLINE=97405699; PubMed=9260743;  
RA Ma D., Nothias F., Boyne L.J., Fischer I.;  
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)  
in rat CNS and PNS during development.";  
RL J. Neurosci. Res. 49:319-332(1997).  
CC -1- FUNCTION: The function of brain MAPs is essentially unknown.  
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes  
CC that accompany neurite extension. Possibly MAP1B binds to at least  
CC two tubulin subunits in the polymer, and this bridging of subunits  
CC might be involved in nucleating microtubule polymerization and in  
CC stabilizing microtubules.  
CC -1- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate  
CC with MAP1B and MAP1B proteins.  
CC -1- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,  
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,  
CC heart or muscle.  
CC -1- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic  
CC nerve levels are high early in development but decrease during  
CC postnatal development and are low in adults. In dorsal root  
CC ganglia levels remain high throughout development.  
CC -1- INDUCTION: By nerve growth factor.  
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
CC KKER and KKEI/V, repeated but not at fixed intervals, which is  
CC responsible for the binding of MAP1B to microtubules.  
CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated  
CC from MAP1B by proteolytic processing. It is free to associate with  
CC both MAP1A and MAP1B. It interacts with the amino-terminal region

CC of MAP1B (By similarity).  
CC -1- PTM: Phosphorylated.  
CC -1- SIMILARITY: TO MAP1A.  
CC -1- CAUTION: A C-terminal fragment of this protein (residues 1597 to  
CC 2459) was originally described as neuraxin in Ref.3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U52950; AAB17068.1; -;  
DR EMBL: X60370; CAC16162.1; -;  
DR EMBL: X16623; CAA34620.1; ALT\_SEQ.  
DR PIR: A56577; A56577.  
DR InterPro: IPR000102; MAP1B\_neuraxin.  
DR Pfam: PF00414; MAP1B\_neuraxin: 10.  
DR PROSITE: P800230; MAP1B\_NEURAXIN; 8.  
KW Microtubules; Repeat; Phosphorylation.  
FT CHAIN ? 2459 MAP1 LIGHT CHAIN LC1.  
FT REPEAT 1869 1885 MAP1B 1.  
FT REPEAT 1886 1902 MAP1B 2.  
FT REPEAT 1903 1919 MAP1B 3.  
FT REPEAT 1920 1936 MAP1B 4.  
FT REPEAT 1937 1953 MAP1B 5.  
FT REPEAT 1954 1970 MAP1B 6.  
FT REPEAT 1988 2004 MAP1B 7.  
FT REPEAT 2005 2021 MAP1B 8.  
FT REPEAT 2022 2038 MAP1B 9.  
FT REPEAT 2039 2055 MAP1B 10.  
FT DOMAIN 559 1035 GLU-RICH.  
FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
FT LYS-AND KKEI/V REPEATS).  
FT LYS-RICH.  
FT DOMAIN 2224 2312 LYS-RICH.  
FT CONFLICT 127 127 M -> V (IN REF. 1).  
FT CONFLICT 140 140 T -> S (IN REF. 1).  
FT CONFLICT 212 212 R -> K (IN REF. 3).  
FT CONFLICT 2169 2169 L -> I (IN REF. 3).  
SQ SEQUENCE 2459 AA; 269497 MW; 2E3F6872DEDB8BA2 CRC64;  
Query Match 10.3%; Score 82.5; DB 1; Length 2459;  
Best Local Similarity 25.4%; Pred. No. 1.9e+02;  
Matches 35; Conservative 25; Mismatches 49; Indels 29; Gaps 6;  
QY 20 GENESTIVSEEDFILPYVKGLEKGYQFDGWEISGFEGKDGAGVYINLSKDTFIKPVFK 79  
DB 593 GKVESKPSVTEKE-----VPSKEEGSPVKAFAVAKRATSKP-----KVTXDKVYVKEIK 642  
QY 80 -KIEEKEEENKPTFDVSKKKDNPVNSQINSHRKEDLOREHSQKSDSTKDTATVTL 138  
DB 643 TPPEEKKEE--KPKKEVAKED-----KTPLEKDEKPKKEAKEIKKEIK 686  
QY 139 --DKNNISSKSTNNPNK 154  
DB 687 KEEKEKKEKVEKKEKTPK 704  
RESULT 33  
YMB4\_YEAST STANDARD; PRT; 279 AA.  
AC P49957;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Hypothetical 32.4 kDa protein in TAP40-ERV25 intergenic region.  
GN YMB14W OR YMB571.04.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxId=4932;

```

RN SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagers K., Lye G., Mole S., Odell C., Pearson D., Rajadream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrall B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII."
RL Nature 387:90-93(1997).
CC -1- SIMILARITY: TO S.POMBE SPAC13D6.03C AND SOME, TO C.ELEGANS
CC C1AB1.5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z49810; CAA8938.1; -.
CC PIR; S55105; S55105.
CC SGD; S0004476; YML014W.
CC InterPro; IPR001601; Methyltransf.
CC InterPro; IPR000051; SAM_bind.
CC Hypothetical protein.
CC KW SEQUENCE 279 AA; 32438 MW; 2F8795BB3C640D3B CRC64;
SQ
Query Match 10.3%; Score 82; DB 1; Length 279;
Best Local Similarity 28.9%; Pred. No. 20;
Matches 24; Conservative 14; Mismatches 39; Indels 6; Gaps 3;
QY 51 WEISGEKKDAGYVNLSDRTFKPYFKIEEKEENKPTDVSCKKDPVNHSGQLN 110
DB 145 WALE-QGSSRRGYHEGMEDVFPWVLPKSKSKPKTKSP--AKVKTRPKNLMNIP 199
QY 111 ESHRKEPDLR-EHSHQKSDSTKD 132
DB 200 PKERSEYLORWKEQOQSKSLDD 222
ID YKTS_YEAST STANDARD; PRT; 427 AA.
AC P36046;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypothetical 47.4 kDa protein in PAB1-MST1 intergenic region.
GN YKL195W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z28195; CAA82039.1; -.
CC PIR; S38032; S38032.
CC SGD; S0001678; YKL195W.

```

```

KW Hypothetical protein.
SQ SEQUENCE 427 AA; 47416 MW; 44386D250DE5DED4 CRC64;
Query Match 10.3%; Score 82; DB 1; Length 427;
Best Local Similarity 27.4%; Pred. No. 35;
Matches 29; Conservative 18; Mismatches 43; Indels 16; Gaps 3;
QY 41 ELEKGYQFGWMEISGEKKDAGYVNLSDRTFKPYFKIEEKEENKPTDVSCKKD 100
DB 142 ETEAGQGLGDDXIGASKVADEGLVLAEDN-----KSSDDKOTDESK-----VSTKDD 191
QY 101 NPQVNHSGQLNESHRKEDIQREHSHQKSDSTKDVATVLDKNNISSK 146
DB 192 -----EQSNEDNATYNNQKDENISENSEENTSKTLDNNAGSSSE 231
ID GLNA_PYRFU STANDARD; PRT; 439 AA.
AC 005907;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN GLNA OR PF0450.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=94365840; PubMed=7916055;
RA Brown J.R., Maechli Y., Robb F.T., Doolittle W.F.;
RT "Evolutionary relationships of bacterial and archaeal glutamine
RT synthetase genes."
RL J. Mol. Evol. 38:566-576(1994).
[2]
RP SEQUENCE FROM N.A.
RA STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RC Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L12410; AAA71968.1; -.
CC EMBL; AE010168; AAL80574.1; -.
CC DR HSSP; P06201; ILGR.
CC DR InterPro; IPR001691; GLN_synth.
CC DR InterPro; IPR004809; GLNA.
CC DR InterPro; IPR001637; GLNA_adenyltn.
CC DR Pfam; PF001020; gln-synt; 1.
CC DR Pfam; PF03951; gln-synt_N; 1.
CC DR ProDom; PD001057; Gln_synt_C; 1.
CC DR TIGRPFAMs; TIGR00653; GLNA; 1.
CC DR PROSITE; PS00180; GLNA_1; 1.
CC DR PROSITE; PS00181; GLNA_ATP; 1.
CC KW ligase; Complete proteome.
CC FT BINDING 358
CC FT AMP (UNDER CONDITIONS OF ABUNDANT
CC FT GLUTAMINE) (BY SIMILARITY).
CC FT K -> G (IN REF. 1).
CC FT M -> L (IN REF. 1).
CC FT CONFLICT 203 203
CC FT CONFLICT 232 232

```

FT CONFLICT 351 351 A -> S (IN REF. 1).  
 FT CONFLICT 370 370 I -> L (IN REF. 1).  
 FT CONFLICT 376 376 S -> N (IN REF. 1).  
 FT CONFLICT 382 382 E -> G (IN REF. 1).  
 FT CONFLICT 397 397 A -> S (IN REF. 1).  
 FT CONFLICT 421 426 MPDTR -> IPPDTE (IN REF. 1).  
 SQ SEQUENCE 439 AA; 50183 MW; 96563A8E9A0E0892 CRC64;

Query Match 10.3%; Score 82; DB 1; Length 439;  
 Best Local Similarity 18.1%; Pred. No. 33;  
 Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

QY 45 GYQFGMEISGPEGKKDAGYVNLNLSKDTFI-----KP-----VFK 79  
 DB 45 GISFGSSVPGQIGIEDSLVFKADPTVEVPMVNARVYGYKDKNPKGADPRGILK 104  
 QY 80 KIEEKEEN-----KPTFVSKKKN----- 101  
 DB 105 RALEEKEGYKAYIGPEPEFLFKNGTWLEIPDVGVFDILTLDKARDIRREIAEYM 164  
 QY 102 -----PQVNSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 143  
 DB 165 PSFGILPEVLHHEVGKAQHEIDFRYDEALKTADNIVSFYITKAVAEHMG 215

## RESULT 36

GLNA\_PYRO STANDARD; PRT; 439 AA.

AC P36687;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glutamine synthetase I (EC 6.3.1.2) (Glutamate--ammonia ligase) (GSI).  
 GN GLNA.

OS Pyrococcus woesei.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxId=2262;

RP SEQUENCE FROM N.A.

RA MEDLINE=93259940; PubMed=8098326;  
 RA Tiboni O., Cammarano P., Sanangelantoni A.M.;

RT "Cloning and sequencing of the gene encoding glutamine synthetase I  
 from the archaeum Pyrococcus woesei: anomalous phylogenies inferred  
 from analysis of archaeal and bacterial glutamine synthetase I  
 sequences";

RT J. Bacteriol. 175:2961-2969 (1993).

CC -1- CARBOLYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
 L-glutamine.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, X60161; CAA42730.1; -.

DR PIR; A36911; A36911.

DR HSP; P06201; ILGR.  
 DR InterPro; IPR001691; GLN synth.  
 DR InterPro; IPR004809; GLN.  
 DR InterPro; IPR001637; GLN adenyltn.

DR Pfam; PF03951; gln-synt\_N; 1.  
 DR Pfam; PD001057; gln-synt\_C; 1.  
 DR TIGRFAMs; TIGR00653; GLNA\_1.

DR PROSITE; PS00180; GLNA\_1; 1.  
 DR PROSITE; PS00181; GLNA\_ATP; 1.

KM Ligase.

FT BINDING 358 358 AMP (UNDER CONDITIONS OF ABUNDANT  
 FT SEQUENCE 439 AA; 50066 MW; C503BA3B790BD25C CRC64;  
 SQ SEQUENCE

Query Match 10.3%; Score 82; DB 1; Length 439;  
 Best Local Similarity 18.1%; Pred. No. 33;  
 Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

QY 45 GYQFGMEISGPEGKKDAGYVNLNLSKDTFI-----KP-----VFK 79  
 DB 45 GISFGSSVPGQIGIEDSLVFKADPTVEVPMVNARVYGYKDKNPKGADPRGILK 104  
 QY 80 KIEEKEEN-----KPTFVSKKKN----- 101  
 DB 105 RALEEKEGYKAYIGPEPEFLFKNGTWLEIPDVGVFDILTLDKARDIRREIAEYM 164  
 QY 102 -----PQVNSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 143  
 DB 165 PSFGILPEVLHHEVGKAQHEIDFRYDEALKTADNIVSFYITKAVAEHMG 215

## RESULT 37

CENC\_HUMAN STANDARD; PRT; 943 AA.

AC 003188; OSPOM5;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Centromere protein C (CENP-C) (Centromere autoantigen C).  
 GN CENPC1 OR CENPC.  
 OS Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_TaxId=9606;  
 RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=placenta;  
 RX MEDLINE=92323541; PubMed=1339310;  
 RA Satoh H., Tomkiet J., Cooke C.A., Rattie H. III, Maurer M.,  
 RA Rothfield N.F., Barnshaw W.C.;

RT "CENP-C, an autoantigen in scleroderma, is a component of the human  
 inner kinetochore plate.";  
 RL Cell 70:115-125 (1992).

RN [2]

RP SEQUENCE OF 1-21 FROM N.A.

RA Poppe M., Boltz J., Hahn B., Dobat K., Eickelbaum W., Pawelietz N.,  
 RA Arand M., Knehr M.;

RT "Promoter characterization of centromere protein C reveals its  
 participation in cell cycle regulation in late G1-phase and expression  
 control by E2F-1, pRB, p107 and Sp-1";

RT Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: Component of the inner kinetochore plate. Required for  
 normal kinetochore assembly.

CC -1- SUBUNIT: Binds to DAXX.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, M95724; AAAS1974.1; -.

DR EMBL; AF151723; AAF73191.1; -.

DR PIR; A42681; A42681.  
 DR Genew; HGNC:1854; CENPC1.  
 DR MIM; 117141; -.

DR GO; GO:0005699; C:kinetochore; TAS.  
 DR GO; GO:0003677; F:DNA binding activity; TAS.  
 DR InterPro; IPR007113; Cupin sup.  
 KM Chromosomal protein; Nuclear protein; DNA-binding; Phosphorylation;

```
KM Centromere. 259 273 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 484 499 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 558 574 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 780 798 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD RES 75 75 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 732 732 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 943 AA; 106925 MW; 6D80810A3E476376 CRC64;

Query Match 10.3%; Score 82; DB 1; Length 943;
Best Local Similarity 23.2%; Pred. No. 75;
Matches 44; Conservative 18; Mismatches 76; Indels 52; Gaps 7;

QY 3 GEVSELEKPRHV-IVTJONGKEMSTIVSEEDFLIPYKGLKGYQFDGMEISGFQKXD 61
DB 343 GRKSRREHNIILPKTLANDHSHKHPHETS--QPSDKTVLDTYSALIDETVNNVSTKY 400
QY 62 AGVIVINLSKDTFKIPVKFKEKK-----EENKPTFDVSKKD-NPQVNSQLNESHK 114
DB 401 EWSKSAEKPSRKRTTKQKQKRFMAKPAEQ-----LDVGQKDNINHTSHITQDEPQR 456
QY 115 KEDLQREH-----SQKSDSTKD-----VYAT 136
DB 457 NSDRNMEHEHMGNDVSKQMPVGSKSKSTRKDESKKRFSSSKNKLVEPVTST 516
QY 137 VLDKNNISSK 146
DB 517 VTKSRRISSR 526

RESULT 38
IP2_HELPJ STANDARD; PRT; 949 AA.
ID IP2_HELPJ
AC Q9ZM46;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-2.
GN INFB OR JHP0377.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Thummino P.J., Carnuso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -1- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB001472; AAD05948.1; -.
DR PIR; E71940; E71940.
DR HAMAP; MF_00100; -. 1.
```

```
DR Interpro; IPR000795; EF_GTPbind.
DR Interpro; IPR004161; EFTU_D2.
DR Interpro; IPR000178; IF2.
DR Interpro; IPR006847; IF2_N.
DR Interpro; IPR005225; Small_GTP.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 2.
DR Pfam; PF04760; IF2_N; 1.
DR Pfam; PF04760; IF2_N; 1.
DR Pfam; PF04760; IF2_N; 1.
DR TIGRfam; TIGR00487; IF-2; 1.
DR TIGRfam; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 451 599 G-DOMAIN.
FT NP_BIND 457 464 GTP (BY SIMILARITY).
FT NP_BIND 503 507 GTP (BY SIMILARITY).
FT NP_BIND 557 560 GTP (BY SIMILARITY).
FT DOMAIN 321 326 POLY-GLU.
SQ SEQUENCE 949 AA; 105961 MW; FA8969B0C64B3278 CRC64;

Query Match 10.3%; Score 82; DB 1; Length 949;
Best Local Similarity 28.7%; Pred. No. 75;
Matches 29; Conservative 20; Mismatches 20; Indels 32; Gaps 5;

QY 70 KDTFKIPVKFKEKKKEENKPTFDVSKKD-----NPO-VNHSQNLNES 112
DB 98 EETKQPKPKTKTKKKKEAPAP--IIRKKEIYVNTFENQTPYENTPKAVSHQIEKA 154
QY 113 HRKEDLQREHHSQKSDSTQVATVLDKNNISSKSTNNPN 153
DB 155 -KQKLQEIQKREA-----LNKLTQSNVTNTNAN 183

RESULT 39
RBP2_PLAVB STANDARD; PRT; 1251 AA.
ID RBP2_PLAVB
AC O00759;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 2 (Fragment).
GN RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RA "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites.";
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M88098; AAA29744.1; -.
DR KW Malaria; Receptor; Membrane.
FT NON_TER 1
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 10.3%; Score 82; DB 1; Length 1251;
Best Local Similarity 22.9%; Pred. No. 1e+02;
DR HAMAP; MF_00100; -. 1.
```

Matches	39;	Conservative	35;	Mismatches	66;	Indels	30;	Gaps	8
0y	5	VSELKPH--RTVTYIQNGKEMSSRTI-----VSEEDILTPYKGELEKGYQFDSWEISG	55						
Db	266	LSLEIKYKDKCTTETLSNKRGRKDKLEFEKEPNESSNKNKNIENINENINSQYLDK	325						
0y	56	FE-GKKDAGYAINL--SKDTFIKPYPKFK-----IEEKEEENKPFYDVK--KKONPOV-	104						
Db	326	IEDAEKASTVELFPHKHTTISNIFKSESLIGVTGSKQKIKNAEDIMKEIERNSIQ	385						
0y	105	-----NHSQLESNRHREKDLQREHSQKSDSTYDVTATYLD--KXNIS	144						
Db	386	TGVKGFQGNLKNLKNLNEPHNYDNADELDNDKSTYNAKVLLETNLSEVXNMLS	435						
RESULT 40									
CRT1_ARATH	CRT1_ARATH	STANDARD;	PRT;	425	AA.				
AC	004151;								
DT	16-OCT-2001 (Rel. 40, Created)								
DT	16-OCT-2001 (Rel. 40, Last sequence update)								
DT	28-FEB-2003 (Rel. 41, Last annotation update)								
DE	Calreticulin 1 precursor.								
GN	CRT1 OR AT1G56340 OR F13N6.20 OR F14G9.5.								
OS	Arabidopsis thaliana (Mouse-ear cress).								
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;								
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;								
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.								
OX	NCBI_TaxID=3702;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=97303616; PubMed=9159940;								
RA	Nelson D.E., Glaunsinger B., Bohner H.U.;								
RT	"Abundant accumulation of the calcium-binding molecular chaperone								
RT	calreticulin in specific floral tissues of Arabidopsis thaliana."								
RT	Plant Physiol. 114:29-37(1997).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=cv. Columbia;								
RX	MEDLINE=21016719; PubMed=11130712;								
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,								
RA	White O., Alonso J., Altrafi H., Araujo R., Bowman C.L., Brooks S.Y.,								
RA	Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,								
RA	Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,								
RA	Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,								
RA	Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,								
RA	Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,								
RA	Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,								
RA	Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,								
RA	Liu X., Liu S.X., Liu Z.A., Luros J.S., Malti R., Marziani A.,								
RA	Miltschker J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,								
RA	Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,								
RA	Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,								
RA	Sun H., Tallon L.J., Tambunga G., Tortini M.J., Town C.D.,								
RA	Utebeck F., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,								
RA	Wu D., Xu G., Fraser C.M., Venter J.C., Davis R.W.;								
RT	"Sequence and analysis of chromosome 1 of the plant Arabidopsis								
RT	thaliana."								
RL	Nature 408:816-820(2000).								
CC	-1- FUNCTION: THIS PROTEIN BINDS CALCIUM, THERE ARE BOTH HIGH AND								
CC	LOW AFFINITY CALCIUM-BINDING SITES (BY SIMILARITY).								
CC	-1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).								
CC	-1- SIMILARITY: BELONGS TO THE CALRETICULIN FAMILY.								
CC	-----								
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -								
CC	the European Bioinformatics Institute. There are no restrictions on way								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								

```

DR EMBL; AC058785; AAC51504.1; -
DR EMBL; AC069159; AAG50908.1; -
DR C96605; C96605.
DR SWISS-2DPAGE; 004151; ARATH.
DR InterPro; IPR001580; Calreticulin.
DR InterPro; IPR000886; ER target.
DR Pfam; PF00262; calreticulin; 1.
DR PRINTS; PR00626; CALRETICULIN.
DR ProDom; PD001866; Calreticulin; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR PROSITE; PS00803; CALRETICULIN_1; 1.
DR PROSITE; PS00804; CALRETICULIN_2; 1.
DR PROSITE; PS00805; CALRETICULIN_REPEAT; 2.
DR Endoplasmic reticulum; Calcium-binding; Repeat; Signal; Glycoprotein;
KW Multigene family.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 425 CALRETICULIN 1.
FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 422 425 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 425 AA; 48527 MW; BEC08E2F342642E CRC64;

Query Match 10.2%; Score 81.5; DB 1; length 425;
Best Local Similarity 21.6%; Pred. No. 35;
Matches 32; Conservative 28; Mismatches 63; Indels 25; Gaps 5

OY 7 ELKPRVTVTTIQNGKMSSTIVSEED-----FILPVYK-----GELEKGYQFDGWEIS 54
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 278 EMKPKKIKNPAYKKGKWKAPMIDNDFKDDPELYVFPKLYGVGLWQVSGSLFDNLVLS 337
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

OY 55 GFEGKQDAGYIINISKDTF-----IKPVFKKIEKKKEENKPTFDVSKKKDNPQVNHQ 108
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 338 -----DDPEYAKKLAETWTGKHDAEKKAFADEAEKKREBESK--DAPAESDAEEAEDD 390
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

OY 109 LNESHRKEDLQREHSOKSDSTKOVYAT 136
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB 391 DNEGDDSDNESKSBETKEAETKEAET 418
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Search completed: February 10, 2004, 10:49:53
Job time : 8:15925 secs

```

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 27.2661 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799  
Sequence: 1 DTGEVSEIKPRVTVTTONG.....ATVLDKNNISKSTNNPNK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL.23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_plant:\*
- 10: sp\_rodent:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriopl:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	799	100.0	2119	2	Q9AHT5 streptococc
2	799	100.0	2140	16	Q97RY6 streptococc
3	796	99.6	2144	16	Q8DDP7 streptococc
4	795	99.5	2144	16	Q9E4M8 streptococc
5	118	14.8	775	16	Q8CPK8 streptococc
6	112.5	14.1	361	5	Q9SP15 plasmodium
7	112.5	14.1	379	5	Q9U6C4 plasmodium
8	111.5	14.0	379	5	Q25706 plasmodium
9	110.5	13.8	379	5	Q25705 plasmodium
10	110	13.8	346	5	Q9U0C0 plasmodium
11	109	13.6	380	5	Q26019 plasmodium
12	109	13.6	3008	5	Q81436 plasmodium
13	108.5	13.6	600	5	Q77355 plasmodium
14	108.5	13.6	1038	13	Q90784 gallus galli
15	108	13.5	354	5	Q25995 plasmodium
16	108	13.5	354	5	Q81J55 plasmodium

17	107.5	13.5	470	10	Q9FJK9	Q9fik9 arabidopsis
18	107	13.4	829	5	Q815F3	Q815f3 plasmodium
19	105	13.1	951	5	Q96229	Q96229 plasmodium
20	103.5	13.0	325	5	Q44016	Q44016 dictyosteli
21	102.5	12.8	2081	10	Q9LH98	Q9Lh98 arabidopsis
22	101.5	12.7	382	5	Q9V7J0	Q9v7j0 drosophila
23	101.5	12.7	556	5	Q9V7J9	Q9v7j9 drosophila
24	101.5	12.7	785	5	Q9G082	Q9g082 drosophila
25	101	12.6	329	5	Q9NFV9	Q9nfv9 plasmodium
26	100	12.5	312	16	Q9PPL5	Q9ppl5 campylobact
27	100	12.5	375	4	Q14712	Q14712 homo sapien
28	100	12.5	1130	5	Q81JZ4	Q81jz4 plasmodium
29	99	12.4	211	5	P91488	P91488 caenorhabdi
30	99	12.4	1859	5	Q81C27	Q81c27 plasmodium
31	99	12.4	2563	5	Q813A0	Q813a0 plasmodium
32	98.5	12.3	238	5	Q81Z26	Q81z26 plasmodium
33	98	12.3	157	5	Q9VOV0	Q9vov0 drosophila
34	98	12.3	4524	5	Q81J39	Q81j39 plasmodium
35	97.5	12.2	449	5	Q81HW3	Q81hw3 plasmodium
36	97.5	12.2	556	5	Q95S93	Q95s93 drosophila
37	97.5	12.2	614	16	Q9K5S1	Q9k5s1 bacillus ha
38	97	12.1	385	5	Q93424	Q93424 caenorhabdi
39	97	12.1	558	5	Q815T1	Q815t1 plasmodium
40	97	12.1	988	10	Q9STN4	Q9stn4 arabidopsis
41	97	12.1	3127	5	Q81DA0	Q81da0 plasmodium
42	96.5	12.1	540	10	Q94C59	Q94c59 arabidopsis
43	96.5	12.1	540	10	Q9S484	Q9s484 arabidopsis
44	96	12.0	421	16	Q8EJQ7	Q8ejq7 shewanella
45	96	12.0	531	5	Q81426	Q81426 plasmodium

## ALIGNMENTS

## RESULT 1

Q9AHT5 PRELIMINARY: PRT; 2119 AA.

AC Q9AHT5: 01-UN-2001 (TRENBLrel. 17, Created)  
DT 01-UN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Serine protease (Fragment).  
GN PRTA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
RA Witzemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
RA Gayle A., Brewah Y.A., Waleh W., Barren P., Lathigra R., Hanson M.,  
RA Langermann S., Johnson S., Koenig S.;  
RT "Use of a whole Genome Approach to Identify Vaccine Molecules  
RT Affording Protection against Streptococcus pneumoniae Infection.";  
RT Infect. Immun. 69:1593-1598(2001).  
RT -I- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
CC AN AMIDE BOND (BY SIMILARITY).  
EMBL AF291699; AAK19159.1; -  
DR HSRP; P00782; 2SRT.  
DR MEROPS; S08.064; -  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRPFMs; TIGR01167; LPXTG\_anchor; 1.

```

DR PROSITE; PS09847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS09840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
  Cell wall; Peptidoglycan-anchor; Protease.
FT NON TER
SQ SEQUENCE 2119 AA; 238226 MW; 517F9B7F6B960A6A CRC64;

Query Match
Best Local Similarity 100.0%; Score 799; DB 2; Length 2119;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGK 60
DB 1932 DTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGK 1991

QY 61 DAGVYINLSKDTPIKVPFKKIEEKEEENKPTFDVSKKKDNPQVNHSQLNESHREKEDLOR 120
DB 1992 DAGVYINLSKDTPIKVPFKKIEEKEEENKPTFDVSKKKDNPQVNHSQLNESHREKEDLOR 2051

QY 121 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 154
DB 2052 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 2085

RESULT 2
Q97RY6 PRELIMINARY; PRT; 2140 AA.
ID Q97RY6
AC Q97RY6;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
DE Serine protease, subtilase family.
GN SP0641.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1313;
RX MEDLINE=21357209; PubMed=11463916;
RA Testelin H., Nelson K.E., Paulsen I.T., Eissen J.A., Read T.D.,
  Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
  Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
  Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
  Holtaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
  McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
  Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
  Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
  "Complete genome sequence of a virulent isolate of Streptococcus
  pneumoniae";
  Science 293:498-506 (2001).
RT EMBL; AE007373; AAK74791.1; -.
DR MEROPS; S08_064; -.
DR TIGR; SP0641; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PRO0723; SUBTILISIN.
DR TIGRPFAM; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS09847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS09840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
  Protease; Complete proteome.

```

```

SQ SEQUENCE 2140 AA; 240426 MW; FA44ADBE2938B334 CRC64;

Query Match
Best Local Similarity 100.0%; Score 799; DB 16; Length 2140;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGK 60
DB 1953 DTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGK 2012

QY 61 DAGVYINLSKDTPIKVPFKKIEEKEEENKPTFDVSKKKDNPQVNHSQLNESHREKEDLOR 120
DB 2013 DAGVYINLSKDTPIKVPFKKIEEKEEENKPTFDVSKKKDNPQVNHSQLNESHREKEDLOR 2072

QY 121 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 154
DB 2073 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 2106

RESULT 3
Q8DOP7 PRELIMINARY; PRT; 2144 AA.
ID Q8DOP7
AC Q8DOP7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
DE Cell wall-associated serine proteinase Prta (EC 3.4.21.-).
GN PRta OR SP0561.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=171101;
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E., Jr., Arnold J., Blaszcak L.C., Burgett S.,
  DeHoff B.S., Bateman S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
  Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
  LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
  McAnen S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
  Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rokey P.,
  Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
  Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R., Jr., Skatrud P.L.,
  Glas J.I.;
  "Genome of the bacterium Streptococcus pneumoniae strain R6";
  J. Bacteriol. 183:5709-5717 (2001).
RT EMBL; AE008434; AAK93365.1; -.
DR KMWI; AE008434; AAK93365.1; -.
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match
Best Local Similarity 99.4%; Score 796; DB 16; Length 2144;
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGK 60
DB 1957 DTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGK 2016

QY 61 DAGVYINLSKDTPIKVPFKKIEEKEEENKPTFDVSKKKDNPQVNHSQLNESHREKEDLOR 120
DB 2017 DAGVYINLSKDTPIKVPFKKIEEKEEENKPTFDVSKKKDNPQVNHSQLNESHREKEDLOR 2076

QY 121 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 154
DB 2077 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 2110

RESULT 4
Q9S4M8 PRELIMINARY; PRT; 2144 AA.
ID Q9S4M8
AC Q9S4M8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)

```

DT 01-MAR-2003 (TRENBLrel. 23, last annotation update)  
 DE Cell wall-associated serine proteinase Prta precursor.  
 GN PRta.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxId=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3.B.  
 RA Bethe G., ten Thoren E., Bongaerts R.J.M., Heinz H.-P., Zysk G.;  
 RT "Cloning and sequencing of a novel surface protease of Streptococcus pneumoniae."  
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL. PEPTIDOLYCAN BY  
 CC AN AMIDE BOND (BY SIMILARITY).  
 CC EMBL; AF127143; AAD4839.1; -.  
 CC DR HSP; P00782; 2SPT.  
 DR MEROPS; S08.064; -.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR006192; LPTTG.  
 DR InterPro; IPR003137; PA.  
 DR InterPro; IPR000209; Peptidase\_S8.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF02225; PA; 1.  
 DR Pfam; PF00082; Peptidase\_S8; 2.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRfams; TIGR01167; LPTG anchor; 1.  
 DR PROSITE; PS08447; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS00840; PA; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 DR Cell wall; Peptidoglycan-anchor; Signal.  
 FT SIGNAL 1  
 FT CHAIN 20 2144  
 FT POTENTIAL.  
 FT CELL WALL-ASSOCIATED SERINE PROTEINASE  
 FT PRta.  
 SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
 Query Match 99.5%; Score 795; DB 2; Length 2144;  
 Best Local Similarity 98.7%; Pred. No. 4,2e-51;  
 Matches 152; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTGEVSELKPHRVVTYVTONGKEMSSITVSEEDFILPVYKGELEKGYQFGWEISGEFGK 60  
 DB 1957 DTGEVSELKPHRITVTYVTONGKEMSSITVSEEDFILPVYKGELEKGYQFGWEISGEFGK 2016  
 QY 61 DAGYINLSKDTFIRPVFKKIEKKKEENKPTFDVSKKKDNQVNHSQLNESHKEDLOR 120  
 DB 2017 DAGYINLSKDTFIRPVFKKIEKKKEENKPTFDVSKKKDNQVNHSQLNESHKEDLOR 2076  
 QY 121 BEHSOKSDSTKDVATATVLDKNNISSKSTNNPNK 154  
 DB 2077 EDHSOKSDSTKDVATATVLDKNNISSKSTNNPNK 2110  
 RESULT 5  
 OSCPK8 PRELIMINARY; PRT; 775 AA.  
 ID OSCPK8  
 AC OSCPK8; 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, last annotation update)  
 DE Penicillin-binding protein 1.  
 GN SE0856.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxId=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,

RA Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AE016746; AAO04453.1; -.  
 KM Complete proteome.  
 SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;  
 Query Match 14.8%; Score 118; DB 16; Length 775;  
 Best Local Similarity 27.2%; Pred. No. 0.69;  
 Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
 QY 1 DTGEVSELKPHRVVTYVTONGKEMSSITVSEEDFILPVYK-----GELEKGYQFGW--- 51  
 DB 615 DSVNAQSILKP-----ITIGNGKQIKQOVSYSSTKVLPHSKWMLMTDGLTMP-DMGWTKE 669  
 QY 52 EISGE-----GKKDGYVIN--LSKDTIRKPVFKKIEKKKEENKPTFDVS---KK 98  
 DB 670 DVLAFEDLTIKIKVSTKNGFVYNOSISKQITK-----NKKIEVLSAEDT 716  
 QY 99 KDNPVNHSQLNESHKEDLQREHSOKSDSTKDVATATVLDKNNISSKSTNN 151  
 DB 717 DDDQKTEBDSDNKSKDKADHDHNTSSSTKN-----DKSNADSDSD 763  
 RESULT 6  
 OSCPK8 PRELIMINARY; PRT; 361 AA.  
 ID OSCPK8  
 AC OSCPK8; 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, last annotation update)  
 DE Merozoite surface protein 3 (Fragment).  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxId=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVO.  
 RA Hiaseda H., Saul A., Long C.A., Miller L.H., Stowers A.W.;  
 RT "Merozoite Surface Protein 3 and Protection Against Malaria in Aotus Monkeys."  
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AY044180; AAK94780.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 361  
 FT SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;  
 Query Match 14.1%; Score 112.5; DB 5; Length 361;  
 Best Local Similarity 22.4%; Pred. No. 0.78;  
 Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8;  
 QY 1 DTGEVSELKPHRVVTYVTONGKEMSSITVSEEDFILPVYKGELEKGYQ----- 47  
 DB 147 ETGE-----ENSRNNFYTTKTE-----YAGKVEKDYERAKANAYOKANQAV 187  
 QY 48 -----FD-----GWEISGF--EGKKDAG-----YVINLSKDTFIRPVFKKIEKKKEEN 89  
 DB 188 LKAKEASSDYILIGWFGGVPEHKKENMLSHLYVSSKDKENMSKENDVDVLE-KEEBA 246  
 QY 90 KPTFDVSKKKDNQVNHSQLNESHKEDLQREHSOKSDSTKDVATATVLDKX----- 141  
 DB 247 ESTEEELKKEENKPTFDVS-----EEDDEEEEEEKKEENKPTFDVS----- 306  
 QY 142 --NISKSTNN 151  
 DB 307 AQNLSKQNNN 318  
 RESULT 7  
 OSCPK8 PRELIMINARY; PRT; 379 AA.  
 ID OSCPK8  
 AC OSCPK8; 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, last annotation update)  
DE Polymorphic antigen.  
GN MSP-3.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FCJ/HN;  
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;  
RT "Sequence of Plasmodium falciparum secreted polymorphic antigen  
gene."  
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AF186190; AAF04099.1; -  
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1F9D5F25 CRC64;  
Query Match 14.1%; Score 112.5; DB 5; Length 379;  
Best Local Similarity 23.6%; Pred. No. 0.82;  
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;  
QY 9 KPHRVVTIÖNGKEMSSIVSEDF-----ILPYKGELEKGYQFD-GWEISGF- 56  
DB 171 KSRINLFSRKTEVAQV--EKDYERAKNAVQKANOAVLKAKKASSYDILGMEFGGV 228  
QY 57 -EGKQDAG-----YVINLSKDTFIKPVFKIEKEEENKPTDVSKKQDPVNHSQLN 110  
DB 229 PEHKKEENMLSHLYVSSKQKENSNDVLDL-KEEBAEETEEELKEEETESIS 287  
QY 111 -----ESHKEDLÖREHHSQKSDSTKYDTATVLDKNNISSKSTNN 151  
DB 288 EDEEEEEEKEEENEKKEQKQKQENNDQKQMEA-----QNLISKQNNN 336  
RESULT 8  
Q25706 PRELIMINARY; PRT; 379 AA.  
AC Q25706;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)  
DT 01-OCT-2002 (TReMBLrel. 22, last annotation update)  
DE Polymorphic antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CS12;  
RX MEDLINE=98156743; PubMed=9497029;  
RA McColl D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the  
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31(1997).  
DR EMBL; U08852; AAC4783.1; -  
SQ SEQUENCE 379 AA; 43302 MW; ABF9D54E1ED91A24 CRC64;  
Query Match 14.0%; Score 111.5; DB 5; Length 379;  
Best Local Similarity 23.6%; Pred. No. 0.98;  
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;  
QY 9 KPHRVVTIÖNGKEMSSIVSEDF-----ILPYKGELEKGYQFD-GWEISGF- 56  
DB 171 KSRINLFSRKTEVAQV--EKDYERAKNAVQKANOAVLKAKKASSYDILGMEFGGV 228  
QY 57 -EGKQDAG-----YVINLSKDTFIKPVFKIEKEEENKPTDVSKKQDPVNHSQLN 110  
DB 229 PEHKKEENMLSHLYVSSKQKENSNDVLDL-KEEBAEETEEELKEEETESIS 287  
QY 111 -----ESHKEDLÖREHHSQKSDSTKYDTATVLDKNNISSKSTNN 151  
DB 288 EDEEEEEEKEEENEKKEQKQKQENNDQKQMEA-----QNLISKQNNN 336  
RESULT 9

Q25705 PRELIMINARY; PRT; 379 AA.  
AC Q25705;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, last annotation update)  
DT 01-OCT-2002 (TReMBLrel. 22, last annotation update)  
DE Polymorphic antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1;  
RX MEDLINE=98156743; PubMed=9497029;  
RA McColl D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the  
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31(1997).  
DR EMBL; U08851; AAC4783.1; -  
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;  
Query Match 13.8%; Score 110.5; DB 5; Length 379;  
Best Local Similarity 22.1%; Pred. No. 1.2;  
Matches 43; Conservative 35; Mismatches 48; Indels 69; Gaps 9;  
QY 1 DTGEVSELKPHRVVTIÖNGKEMSSIVSEDFILPYKGELEKGYQ----- 47  
DB 167 ETGE-----RNSRNNFTYTKKE-----YAGKYERAKNAVQKANOAV 207  
QY 48 -----FD-----GWEISGF-EGKQDAG-----YVINLSKDTFIKPVFKIEKEEEN 89  
DB 208 LKAKEASDYDILGMEFGGVPEHKKEENMLSHLYVSSKQKENSNDVLDL-KEEBA 266  
QY 90 KPTFDVSKKQDPVNHSQLN-----ESHKEDLÖREHHSQKSDSTKYDTAT 136  
DB 267 EETEEELKEEENEETESISDEDEEEEEEKEEENEKKEQKQKQENNDQKQMEA- 325  
QY 137 VLDKNNISSKSTNN 151  
DB 326 -----QNLISKQNNN 336  
RESULT 10  
Q25706 PRELIMINARY; PRT; 346 AA.  
AC Q25706;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)  
DT 01-OCT-2002 (TReMBLrel. 22, last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
GN MSP3.  
OS Plasmodium reichenowii.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5854;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20416497; PubMed=10960178;  
RA Okenu D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
RT Plasmodium reichenowii and Plasmodium falciparum.";  
RL Mol. Biochem. Parasitol. 109:185-188(2000).  
DR EMBL; AJ252286; CAB65754.1; -  
FT NON\_TER 1 1  
FT NON\_TER 346 346  
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAF010 CRC64;  
Query Match 13.8%; Score 110; DB 5; Length 346;  
Best Local Similarity 26.2%; Pred. No. 1.1;  
Matches 34; Conservative 26; Mismatches 54; Indels 16; Gaps 5;  
QY 37 VKKGELEKGYQFD-GWEISGF-EGKQDAG-----YVINLSKDTFIKPVFKIEKEEEN 88  
DB 184 VLKAKEASSYNYILGMEFGGVPEHKKEENMLSHLYVSSKQKENSNDVLDL-KEEBA 242

QY 89 NKPTDVSKKKDNPOVNHSHKRDLOREHSOKSDSTKDYATV-----LQKN 141  
 Db 243 AAEETGEQLEKNEKEETSEIINEDEQEBEKEEKEENKKEQAKESQNDQKEDMAQ 302  
 QY 142 NISSKSTNN 151  
 Db 303 NLISKNNNN 312

RESULT 11  
 Q26019 PRELIMINARY; PRT; 380 AA.  
 ID Q26019  
 AC Q26019  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Polymorphic antigen precursor.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FC27;  
 RX MEDLINE=95198774; PubMed=7891748;  
 RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,  
 RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.,  
 RA "Molecular variation in a novel polymorphic antigen associated with  
 RT Plasmodium falciparum merozoites.";  
 RL Mol. Biochem. Parasitol. 68:53-67(1994).  
 DR EMBL; L07944; AAC09378.1; -.  
 KW SIGNAL.  
 FT CHAIN 1 25 POTENTIAL.  
 FT SIGNAL 26 380 POLYMORPHIC ANTIGEN.  
 SQ SEQUENCE 380 AA; 43290 MW; 0986CA1393094CA2 CRC64;

Query Match 13.6%; Score 109; DB 5; Length 380;  
 Best Local Similarity 23.5%; Pred. No. 1.5;  
 Matches 42; Conservative 34; Mismatches 67; Indels 36; Gaps 8;

QY 2 TGEVSELKRVTVITIQNGKMSSTIVSEEDFILPVY-----KGELEKGYQD-CWE 52  
 Db 166 TSETPE-KPSRINLPSRKTEYAEKAKAYEKAQAVLKAKEASSYDYLQWE 224  
 QY 53 ISGF--EGKQAG-----VINLSKDTFKPVFKKIEEKEENKPTFVSKKNPQVN 105  
 Db 225 FGGGVPHEKKENMLSHLYSSKDKENIKENDVDLDE-KEEBAETEEBELEKKEEET 283

QY 106 HSQLN-----ESHRRKEDLOREHSOKSDSTDYATVLDKNNISSKSTNN 151  
 Db 284 ESEISDEEEEBEKEEENKKEQEKESQNSENNNDQKQDWEA-----QNLISKNNNN 337

RESULT 12  
 ID Q81436 PRELIMINARY; PRT; 3008 AA.  
 AC Q81436  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 OS PFE0325W.  
 GN Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=36329;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.,  
 RL Submitted (Sep-2002) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22255708; PubMed=12368867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark E., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhornou A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Ruter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
 RL Nature 419:527-531(2002).  
 DR EMBL; AL929351; CAD51431.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 3008 AA; 356023 MW; 60CBEBE15C59984 CRC64;

Query Match 13.6%; Score 109; DB 5; Length 3008;  
 Best Local Similarity 32.4%; Pred. No. 14;  
 Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

QY 64 YVINLSK---DFFIKVPFKIEEKEENKPTDVSKKKKNPOVNHQ---LNSHKE 116  
 Db 2310 YDIELSKIEKFGASIGPVFT-EENKKEENK--EVNKKENKEENKKEENKKE 2366

QY 117 DLQREH-----SOKSDSTKDYATVLDKNNISSK-----STNNPNK 154  
 Db 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENK 2414

RESULT 13  
 ID 077355 PRELIMINARY; PRT; 600 AA.  
 AC 077355;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Hypothetical 71.7 kDa protein.  
 GN PFC0465C, MAL3P4.20.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=99376085; PubMed=10448855;  
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Felwell T.,  
 RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
 RA Horrocks P., Jagels K., Jasal B., Kyes S., McLean J., Moule S.,  
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,  
 RA Ruter S., Skelton J., Squares R., Squares S., Sulston J.E.,  
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;  
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium  
 RT falciparum.";  
 RL Nature 400:532-538(1999).  
 RL EMBL; AL008970; CA15610.2; -.  
 DR InterPro; IPR002483; PMI.  
 DR Pfam; PF01480; PMI; 1.  
 DR SMART; SM00311; PMI; 1.  
 KW Hypothetical protein.

50	SEQUENCE	600 AA;	71663 MW;	57EAB42565CAD64C	CRC64;	
	Query Match	13.6%;	Score 108.5;	DB 5;	Length 600;	
	Best Local Similarity	29.3%;	Pred. No. 2.7;			
	Matches	54;	Mismatches	22;	Indels 61; Gaps 12.	
Qy	27	IVSEEDFILPYV----	KGLEKGYQPDGWEISGEKK--	DAGYVINSKDTFIPV	77	
Db	60	ILGFEDDILVEYCSIQLKQSEKK--	DGEEDKYNAKKLKITLGTGIGNKKSDIFIEL		116	
Qy	78	FKKI--EEKKEE-----	ENKPEFDSVK--KQDNQVHNSQINE-----	SHRX	115	
Db	117	LELLINEKKKEHIDLTIANENK--	INDIKKYKNENENINENYNEKSDISNCKDEKHSION		175	
Qy	116	E-----	DLQREEH-----	SQKSDSTK-----	DVTATVLDKNNISKSSTTN	150
Db	176	EHNINNVNMLKKEKEYTIDIQDRKRKHRSLSQSDSYKTRPFKRYKTSIER--	SLSNKRYDE		234	
Qy	151	NPNK	154			
Db	235	KTNK	238			

RESULT 14

ID	Q90784	PRELIMINARY;	PRT; 1038 AA.
AC	Q90784;		
DT	01-NOV-1996	(TREMBLrel. 01. Created)	
DT	01-NOV-1996	(TREMBLrel. 01. Last sequence update)	
DT	01-OCT-2002	(TREMBLrel. 22. Last annotation update)	
DE	Claustrin.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
CC	Gallus.		
OX	NCBI_TextID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RA	Cole G.J.;		
RL	Submitted (OCT-1992) to the EMBL/GenBank/DDBJ databases.		
RN	[2]		
RP	SEQUENCE OF 1-451 FROM N.A.		
RC	TISSUE=Brain;		
FX	MEDLINE=94157526; PubMed=7906711;		
RA	Burg M.A., Cole G.J.;		
RT	"Clastrin, an antichedhesive neural keratan sulfate proteoglycan, is		
RT	structurally related to MMPB."		
J	J. Neurobiol. 25:1-32(1994).		
RL	EMBL; X67778; CAA47988.1; "-"		
QO	SEQUENCE 1038 AA; 117112 MW; 213D694A5E510927 CRC64;		

## Query M

Query Match	13.6%;	Score 108.5;	DB 13;	Length 1038;
Best Local Similarity	28.6%;	Pred. No. 4.9;		
Matches	34;	Mismatches	43;	Indels 17;
				Gaps 4.

[illegible]

## RESULT 15

ID	Q25995	PRELIMINARY;	PRT;	354 AA.
AC	Q25995.1			
DT	01-NOV-1996	(TREMBLrel. 01,	Created)	
DT	01-NOV-1996	(TREMBLrel. 01,	Last sequence update)	
DT	01-OCT-2002	(TREMBLrel. 22,	Last annotation update)	
DE	Antigen.			

OS Plasmodium falciparum (Isolate NF54).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=5843;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=NF54;  
RC MEDLINE=95198774; PubMed=7891748;  
RX McCall D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,  
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;  
RT "Molecular variation in a novel polymorphic antigen associated with  
RT Plasmodium falciparum mezozoites.";  
RL Mol. Biochem. Parasitol. 68:53-67(1994).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=NF54;  
RC MEDLINE=98156743; PubMed=9497029;  
RX McCall D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the  
RT Plasmodium falciparum mezozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31(1997).  
DR EMBL: L28825; AAC09377.1; -  
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F46B527 CRC64;

### Query M

Query Match	13.5%;	Score 108;	DB 5;	Length 354;
Best Local Similarity	23.2%;	Pred. No. 1.7;		
Matches	36;	Conservative	29;	Mismatches 46;
				Indels 44;
				Gaps 6

```

OY 1 IONGKMSSTIASBEEFLIPYKGBLEKGYOFDGMELSGF--EKKQDAG-----YVNLIS 69
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 VLKAKRASS-----XDYIL-----CWFEGGQVBEHKKEENMLSHLYVSSKD 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 70 KDFIFIPVPRKIEBKKEE-----ENKPTFVSKKKQNPVHNSQJNSRRE 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 KENISKENDVDLDEKKEBEAEETBEBELEKKNKEEETBSLSDBEBEBEBEKEBEKENDKK 261
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 117 DLQREHSQKSDSTYDVTATVLDKNNISKSTNN 151
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 EOEKEQSNENDQKCDMEA-----ONLISKQNNN 311
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 16

ID	Q81U55	PRELIMINARY;	PRT;	354 AA.
AC	Q81U55;			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	Merozoite surface protein 3.			
GN	PI10_0345.			
OS	Plasmodium falciparum (isolate 3D7).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxId=36329;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=3D7;			
RX	MEDLINE=22255705; PubMed=12368864;			
RA	Gardner M.J., Hall N., Fung E., White O., Berrihan M., Hyman R.W.,			
RA	Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.			
RA	Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,			
RA	Chan M.-S., Nene V., Shallow S.-J., Sun B., Peterson U., Angiuoli S.,			
RA	Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,			
RA	Martin D.M.A., Fairhead A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,			
RA	McGadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,			
RA	Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,			
RA	Fraser C.W., Barrett B.;			
RT	"Genome sequence of the human malaria parasite Plasmodium			
RT	falciparum".			
RL	Nature 419:498-511(2002).			
SD	EMBL, AE014834, AAN35542.1, -. SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;			

Best Lo

Query Match	13.5%;	Score 108;	DB 5;	Length 354
Best Local Similarity	23.2%;	Pred. No. 1.7;		

Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 17 IONGKMSSTIVSEDFILPYKGELEKGYQPDGWEISGF--EGKKDAG----YVNLIS 69  
 DB 181 YKAKEASS-----VDYIL-----GMEFGGVEHKKKEENMLSHLYVSSKD 221  
 QY 70 KOTFIKPVFKKIEKKEE-----ENKPFVDSKKKDNQVNHSLNSHRE 116  
 DB 222 KENISKENDVDLDEKEEAEETBEERLEKNEEETSESLSEDEEEEEEKKEENDKK 281  
 QY 117 DLQREHSQKSDSTKDVATVLDKNNISSKSTNN 151  
 DB 282 EOEKQSNENNOKKDMA-----QNLISKNNNN 311

RESULT 17  
 Q9FJK9 PRELIMINARY; PRT; 470 AA.  
 AC 09FJK9;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
 DE GblAAF20218.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsiis.  
 OC NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=99087489; PubMed=9872454;  
 RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
 RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
 RT physically assigned P1 and TAC clones.";  
 RL DNA Rep. 5:297-308 (1998).  
 RL EMBL; AB015468; BAB10694.1; -  
 SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72B35AC54 CRC64;

Query Match 13.4%; Score 107.5; DB 10; Length 470;  
 Best Local Similarity 20.1%; Pred. No. 2.5;  
 Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 9 KPHRYVTIIONGKMSSTIVSEDFILPYKGELEKGYQPDGWEISGF-----GKK 60  
 DB 82 RENRYTDTYQNNNSNGSK-----YVODLARIYDE-EATGSGSAQRIDHPNK 129  
 QY 61 DAGVYINLSKOTFIKPVFKKIEKKKEENKPTFDVSKKKDN----- 101  
 DB 130 NNGITEKAPENSPIEETSHRVDDNKRINNQKFTAKSSENAVSRVPGADHRAEVMGK 189  
 QY 102 PQVNHSLNE-----SHREKDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPK 154  
 DB 190 PMENHDQVQTESAEKSHRKEENVYSEKPRDQGVAKTEAKDKRKEKKEKTESINK 248

RESULT 18  
 Q815F3 PRELIMINARY; PRT; 829 AA.  
 AC 0815F3;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PFL1275C.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OC NCBI\_Taxid=36329;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=3D7;

RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Anguilo S.,  
 RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Frazer C.M., Barrrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum.";  
 RL Nature 419:498-511 (2002).  
 DR EMBL; AE014848; AAN36341.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 829 AA; 98816 MW; EF2675E301B2CE93 CRC64;

Query Match 13.4%; Score 107; DB 5; Length 829;  
 Best Local Similarity 24.8%; Pred. No. 5;  
 Matches 36; Conservative 30; Mismatches 43; Indels 36; Gaps 6;

QY 26 TIVSEDFILPYKGELEKGYQPDGWEISGFEGKKDAGYVNLISKOTFIKPVFKKIEBK 85  
 DB 519 TLMTRDVLVDVFPHPYMKKTHLN-----KKETLFNFSLN-----FREIEKVK 561  
 QY 86 EEENKPT-----FDVSKKDN-QVNHSLNSH-RKEDLQREHSQKSDS 129  
 DB 562 RDKKKGTHINNKNDAAEYMYLKIKKKKKNEENNTELNDNSIKKENNKLVHE--DNS 618  
 QY 130 TKDVTATVLDKNNISSKSTNNPNK 154  
 DB 619 LKQEOIITINDKQVIEHTKIYDNQKK 643

RESULT 19  
 Q96229 PRELIMINARY; PRT; 951 AA.  
 AC 096229;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PFB0680W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OC NCBI\_Taxid=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=99021743; PubMed=9804551;  
 RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,  
 RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,  
 RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Petrea M.,  
 RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
 RA Frazer C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
 RT falciparum.";  
 RL Science 282:1126-1132 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Anguilo S.,  
 RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Frazer C.M., Barrrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum.";





RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodok A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hossain N., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laevo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RA Science 287:2185-2195 (2000).  
 RL [12]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banson J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsner V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Matrei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Paclebo J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phapleong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svitskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RA "Sequencing of *Drosophila melanogaster* genome.";  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [13]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochownik S.E., Smith C.D.,  
 RA Tupy J.L., Begeman C., Bertram B., Carlson J.W., Celinker S.E.,  
 RA Clamp N., Drysdale R., Emmert D., Frise B., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Seale S.M.J., Smith E., Shu S., Smutnicki F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RA "Annotation of *Drosophila melanogaster* genome.";  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [14]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL [15]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Flybase;  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RL [16]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-20564328; PubMed-10956655;  
 RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
 RA O'Neill K.T., Pocht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
 RA Friedman P.A.;  
 RA "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved  
 RT Isoform of Asph Missing the Catalytic Domain Share Exons with  
 RT Junction.";  
 RT J. Biol. Chem. 275:39543-39554 (2000).  
 DR EMBL, AE003808; AAF58063.2; -;  
 DR EMBL, AF289494; AAC40807.1; -;  
 DR Flybase; Fgn0034075; Asph.  
 SQ SEQUENCE 382 AA; 43287 MW; 605EC03AEBFC6E8B CRC64;

Query Match 12.7% Score 101.5; DB 5; Length 382;  
 Best Local Similarity 24.5%; Pred. No. 5.5;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 31 EDFILPVYKGEIKGYQPDG-----EISGPEKQAGYI-----NLKPTPIK 75  
 DB 78 EDDLTPLESEFSK-VFDGVDHEDHDDHVDQEPGALDDPHDEHDHDDHDEDE 135  
 QY 76 PVFKIIEKKKEENKPT-----FDVSKKKNPVNHSQNLNESHAKELQBEHSQKSDS 129  
 DB 136 PLTELELELELEEEPEETDEPPADEEYEDDEDEENNA--GENTIADEEEEEEEDND 193  
 QY 130 TKDVTATVLDNNISKST 148  
 DB 194 EGVTAIVATTEATTEAT 212  
 RESULT 23  
 ID Q9V719 PRELIMINARY; PRT; 556 AA.  
 AC Q9V719;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE CG8421 protein.  
 GN ASPH OR CG8421 OR CG18658.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidae; Drosophilidae; Drosophila.  
 CX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Abmayant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.A., Bouck J., Brokstein P., Brotlier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodok A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laevo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

[2]

RP SEQUENCE FROM N.A.

RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Bantz J., An H., Baldwin D., Bantz J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Phouenavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Teector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,

RT "Sequencing of Drosophila melanogaster genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[3]

RP SEQUENCE FROM N.A.

RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,

RA Clamp M., Dysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,

RT "Annotation of Drosophila melanogaster genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[4]

RP SEQUENCE FROM N.A.

RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[5]

RP SEQUENCE FROM N.A.

RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AEO03808; AAF58064.2; -

SQ FlyBase; FBgn0034075; Asph.

DR FlyBase; FBgn0034075; Asph. B420980CDB6C357A CRC64;

SQ SEQUENCE 556 AA; 63144 MW; B420980CDB6C357A CRC64;

Query Match 12.7%; Score 101.5; DB 5; Length 556;

Best Local Similarity 24.5%; Pred. No. 8.3;

Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 31 EDFILPVYKGLKGYQFDGM-----EISGFEKKKAGYVI-----NISKQTFIK 75

DB 78 EDIDTFLSRSFGRK--VFQGWVDEHDEHGHVQSPSGEALDDHDDHDDHDEDEDE 135

QY 76 PVKPKLKEKKKEENKPT-----FDVSKKKDNQVNHVSQLNESHKREDLQREHSQKSDS 129

DB 136 PLTEELLEEELLEEELFEDEDEPADEBEDEDEENNA--GENITADAEELBEEDNDND 193

QY 130 TKDVTATVLDKNNISKST 148

DB 194 EGTVEATVEATTEATTEAT 212

RESULT 24

Q9G082 PRELIMINARY; PRT; 785 AA.

AC Q9G082; 01-MAR-2001 (TEMBLrel. 16, Created)

DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)

DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DR Aspartyl beta-hydroxylase variant 1 (CG8421-PA).

GN ASPH OR CG8421 OR CG18658

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

[1]

RP SEQUENCE FROM N.A.

RA MEDLINE=20564328; PubMed=10956665;

RX Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,

RA O'Neil K.T., Poche R.J., Scully M.S., Hollis J.M., Hollis G.F.,

RA Friedman P.A.,

RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved

RT Isoform of Asph Missing the Catalytic Domain Share Exons with

RT Junction."

RL J. Biol. Chem. 275:39543-39554(2000).

[2]

RP SEQUENCE FROM N.A.

RA STRAIN=Berkely;

RC MEDLINE=20196006; PubMed=10731132;

RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhanderi D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Nussbaum D.R., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nussbaum D.R., Pacleb J.M.,

RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Teector C., Turner R., Venter J.C., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).

[3]

RP SEQUENCE FROM N.A.

RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Bantz J., An H., Baldwin D., Bantz J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Phouenavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Teector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,

RT "Sequencing of Drosophila melanogaster genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[4]

RP SEQUENCE FROM N.A.

RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,

RA Clamp M., Drysdale R., Emmert D., Friese E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Seale S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome."  
 RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Flybase;  
 RU Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF289493; MAG40806.1; -;  
 DR EMBL; AB003808; AAM70947.1; -;  
 DR Flybase; FBgn0034075; AspH.  
 DR InterPro; IPR001440; TPR.  
 DR InterPro; IPR006025; Zn\_MTPase.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1  
 SQ SEQUENCE 785 AA; 89843 MW; 30A8DFCD6836F7F1 CRC64;

Query Match 12.7%; Score 101.5; DB 5; Length 785;  
 Best Local Similarity 24.5%; Pred. No. 12;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 31 EPIFLPVYKGELEKGYQFPGW-----EISGREGKKDAGYV-----NLKSDTFLK 75  
 DB 78 ELDLPPLSRPSRK--VFQGWVDEHDEHDGDPSPGALDDHDDHDDHDEDEDE 135  
 QY 76 PVFKLEEKKEENKPT-----FDVSKKQNPQVNHSQLNSHREKDLQREHSQKSDS 129  
 DB 136 PTEELIEELIEEETDEDEBADEYEDEDEENNA--GNITFEDAEKEEEDNDND 193  
 QY 130 TKDVTATVLDKNNISKST 148  
 DB 194 EGTVEATVEATTEATTEAT 212

RESULT 25

Q9NFV9 PRELIMINARY; PRT; 329 AA.  
 AC Q9NFV9;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Merzoitte surface protein 3 (Fragment).  
 GN MSP3.  
 OS Plasmodium falciparum (Isolate 7G8).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=57266;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=7G8;  
 RX MEDLINE=20416497; PubMed=10960178;  
 RA Okenu D.M.N., Thomas A.W., Conway D.J.;  
 RT "Allelic lineages of the merzoitte surface protein 3 gene in  
 RT Plasmodium reichenowi and Plasmodium falciparum."  
 RL Mol. Biochem. Parasitol. 109:185-188 (2000).  
 DR EMBL; AJ252287; CAB85901.1; -;  
 FT NON\_TER 1  
 FT NON\_TER 329  
 SQ SEQUENCE 329 AA; 36916 MW; C5B045DB5E21A159 CRC64;

Query Match 12.6%; Score 101; DB 5; Length 329;  
 Best Local Similarity 25.3%; Pred. No. 5.1;  
 Matches 38; Conservative 25; Mismatches 47; Indels 40; Gaps 7;

QY 17 IONGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YINIS 69  
 DB 171 VIKAKEAS-----YDYL-----GMFGGAVPEHKKENMLSHLYVSKD 211  
 QY 70 KOTFLKPVKTIKKEEENKPTFDVSKKQNPQVNHSQLNSHREKDLQREHS----- 123

DB 212 KENISKENDVLDK--KEEAEETEELIEEKNQEBTEISEDEBEEKEKEQAKE 270  
 QY 124 --SKSDSTKDVATVLDKNNISKSTNN 151  
 DB 271 QSNENNDDQKXMEIA-----QNLISKNNNN 295

RESULT 26

Q9PPL5 PRELIMINARY; PRT; 312 AA.  
 ID Q9PPL5  
 AC Q9PPL5;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Putative membrane protein.  
 GN CU0692C.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Campylobacteraceae; Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCCT 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,  
 RA Jags K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajadream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrall B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences."  
 RL Nature 403:665-668 (2000).  
 DR EMBL; AL139076; CAB72966.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 312 AA; 37221 MW; 0004FA7836A7A1B8 CRC64;

Query Match 12.5%; Score 100; DB 16; Length 312;  
 Best Local Similarity 25.3%; Pred. No. 5.7; Indels 28; Gaps 8;  
 Matches 41; Conservative 32; Mismatches 61; Indels 28; Gaps 8;

QY 17 IONGKEMSTIVSEEDFILPVYK-----GELEKGYQFDGWEISGREGKKDAGYV 65  
 DB 24 INQKPLDDDLRDEISSDILARRKKTTPKFLLEDEEYESKTKTSNYLKED---L 80  
 QY 66 INL---SDTFLKPVYKTIKKEEENKPT--FDVSKKQNP-----QVNHSQLNSHREK 115  
 DB 81 INVKLEEKQSLAKKIFSKERKREKKNKFLFSRKKANEIKNIQTQIOITKSNQA 140  
 QY 116 EDLQREHSQKSDSTKDV--TATVLDKNNISK--STNNPN 153  
 DB 141 TTQTKQEKELTNSIEKIQTETKIQKPLIEKLDVKNQPN 182

RESULT 27

014712 PRELIMINARY; PRT; 375 AA.  
 ID 014712  
 AC 014712;  
 DT 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 15, Last annotation update)  
 DE Cell cycle progression restoration 8 protein.  
 GN CPR8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98043401; PubMed=9383053;  
 RA Edwards M.C., Liegeois N., Horecka J., Depinho R.A., Sprague G.F. Jr.,  
 RA Tyers M., Blledge S.J.;  
 RT "Human CPR (cell cycle progression restoration) genes impart a far-

RT phenotype on yeast cells."  
RL GenBank: AF011794; AAB69314.1; --  
DR EMBL: AF011794; AAB69314.1; --  
DR InterPro: IPR004238; LEA; 1.  
DR Pfam: PF02987; LEA; 1.  
SQ SEQUENCE 375 AA; 44320 MW; 0C261BF68AAEF7AD CRC64;

Query Match 12.5%; Score 100; DB 4; Length 375;  
Best Local Similarity 28.0%; Pred. No. 7;  
Matches 51; Conservative 16; Mismatches 67; Indels 48; Gaps 9;

QY 3 GEVSEIKPHRVTVTTONGKMSSTIVSE--EDFILPYKGELEKGYQF-----DGE----- 52  
DB 18 GELQOQSGSGL-----HGKSDSPNVYTEKKEIALILBERLTELRLKTLFEQQRSDLWRLY 72  
QY 53 -----ISGEGKCDAG-----YVNLKDTFKIPVFKIEKKEENKPTDVSKKD 100  
DB 73 VEAKDNGKQGTGKGGKGRSHRVNKKSKGTFLGSV-----KETFDAMKNGT 120  
QY 101 NPQVNHSQLNESHKEDLQREHESQKSDST-----KQVATATVLD-KNNISKSTTNNP 152  
DB 121 KEFVRHHEKIKQAKEDV-KENLKPSDSVKTFFHFKOTTNIFDEKGNKRNATKEA 179  
QY 153 NK 154  
DB 180 EK 181

## RESULT 28

Q81J24 PRELIMINARY; PRT; 1130 AA.  
ID 081J24  
AC 081J24  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF10\_0046.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=22255705; PubMed=12368864;  
RA Gardner M.J., Hall N., Pung B., White O., Berriman M., Hyman R.W.,  
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiuoli S.,  
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrell B.;  
RT "Genome sequence of the human malaria parasite Plasmodium  
falciparum";  
RT Nature 419:498-511(2002).  
DR EMBL: AB014629; AAN35244.1; --  
DR Hypothetical protein.  
SQ SEQUENCE 1130 AA; 131697 MW; F66AFAE08C4CCDB CRC64;

Query Match 12.5%; Score 100; DB 5; Length 1130;  
Best Local Similarity 26.1%; Pred. No. 23;  
Matches 30; Conservative 18; Mismatches 53; Indels 14; Gaps 3;  
QY 54 SGFEGKDAAGV--INLSKDTFKIPVFKIEKKEENKPTDVSKKDNPCVNHSQLNE 111  
DB 568 SFLSGNDYRIDRIYISPNVYSNNRNININENAVIGVETFRSDYNSHDETMB 627  
QY 112 SHREKDLQREHESQKSDST-----DSTQVATATVLDKN-NISSKSTTNNPK 154  
DB 628 NNEKEINREEGNEEKRSKRKKTNDYDNKDYVNNAGSGNSHNSNNNNNNK 682

## RESULT 29

P91488 PRELIMINARY; PRT; 211 AA.  
ID P91488  
AC P91488;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical 24.0 kDa protein.  
GN T23B3.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Magdi L., Le T.;  
RT "The sequence of C. elegans cosmid T23B3.";  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U88309; AAB42334.1; --  
DR WormPeP: T23B3.5; CE14016.  
KW Hypothetical protein.  
SQ SEQUENCE 211 AA; 23955 MW; 11B16164A87E5928 CRC64;

Query Match 12.4%; Score 99; DB 5; Length 211;  
Best Local Similarity 29.4%; Pred. No. 44;  
Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

QY 57 EGGKDAAGVYINLSKDTFKIPVFKIEKKEENKPTDVSKKDNPCVNH-----SQLNES 112  
DB 69 EGGKDGEGKSKKDGKDEKKEKDEKDGKDEKDDKDEKDDKDEKDDKDEKDDKDEKDDKDE 128  
QY 113 HKKEDLQREHESQKSDSTQVATATVLDKNNISKSSTTNNPK 154  
DB 129 EKKDKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDE 170

## RESULT 30

Q81C27 PRELIMINARY; PRT; 1859 AA.  
ID 081C27  
AC 081C27;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF07\_0016.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,  
RA Quail M., Barrell B.;  
RT Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL844506; CAD50814.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 1859 AA; 218376 MW; 2266544164BD360C CRC64;

Query Match 12.4%; Score 99; DB 5; Length 1859;

Best Local Similarity 24.7%; Pred. No. 48;  
Matches 36; Conservative 24; Mismatches 60; Indels 26; Gaps 5;

QY 26 TIVSEDFILPVYKGELEKGYOPDMEISGF-----EKKQAGVYINLSKDT- 72  
DB 216 TDYSEDEIYAKYIOKDSKSDNSYOGYDKSKLINTSNINMLNVTNKKQVNH--SMSSTNT 273  
QY 73 -----FIRVFKEIKKEEENKPTFDVSKKQDNQVNHSQLNESHKEDLQREHSQKS 127  
DB 274 QODLSFIHSSINKYKKEKKEK--NYDKKKSNTNDSYNTONDPKKNQKSFVDNN 332  
QY 128 DSTKQVATVLDKNNISSKSTNNPN 153  
DB 333 NKGRND-----HNKNNELEQVYNNPN 353

RESULT 31  
0813A0  
ID 0813A0 PRELIMINARY; PRT; 2563 AA.  
AC 0813A0;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DE E1-E2 ATPase/hydrolase, putative.  
CN PF10240C.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22255708; PubMed=12368867;  
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
RA Felwell T., Goble A., Goodhead I., Gilliam R., Hamlin N., Hance Z.,  
RA Harper D., Hauser H., Hornsby T., Holtroyd S., Horrocks P.,  
RA Humphrey S., Jagels K., Kyes K.D., Johnson D., Kethorou A.,  
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
RA Seeger K., Sharp S., Smith R., Squares R., Stevens K.,  
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
RA Sulten J.E., Craig A., Newbold C., Barrell B.G.  
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."  
RL Nature 419:527-531(2002).  
DR EMBL: AL929355; CAD51734.1; --  
KW Hydrolase.  
SQ SEQUENCE 2563 AA; 298805 MW; 3F9613243D26F8F1 CRC64;

Query Match 12.4%; Score 99; DB 5; Length 2563;  
Best Local Similarity 29.4%; Pred. No. 68;  
Matches 35; Conservative 18; Mismatches 60; Indels 6; Gaps 4;

QY 40 GELKGYOPDG-WEISGFEKQDAGVYIMLSKDTPIKPVFKKIEEKEENKPTFDVSKK 98  
DB 856 GEEKG-NIDGIYILKQKXKXKMDIKGEBEKNDFEKKSEKSDNENSEIDKNNVYLK 914  
QY 99 KD--NPOVNHSQLNESHKEDLQRE-HSQKSDSTKYDTATVLDKNNISSKSTNNPN 153  
DB 915 KEKHNSNDENDINDSVLKEVKEKEIHSSSSNEQSDSYLKKIEKKDKNNISVDNNEN 973

RESULT 32  
081226  
ID 081226 PRELIMINARY; PRT; 238 AA.  
AC 081226;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Hypothetical protein.

GN PF10765M.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22255708; PubMed=12368867;  
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
RA Felwell T., Goble A., Goodhead I., Gilliam R., Hamlin N., Hance Z.,  
RA Harper D., Hauser H., Hornsby T., Holtroyd S., Horrocks P.,  
RA Humphrey S., Jagels K., Kyes K.D., Johnson D., Kethorou A.,  
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
RA Seeger K., Sharp S., Smith R., Squares R., Stevens K.,  
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
RA Sulten J.E., Craig A., Newbold C., Barrell B.G.  
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."  
RL Nature 419:527-531(2002).  
DR EMBL: AL929355; CAD51839.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 238 AA; 28481 MW; DA175A2B5A109B49 CRC64;

Query Match 12.3%; Score 98.5; DB 5; Length 238;  
Best Local Similarity 28.1%; Pred. No. 5.5;  
Matches 48; Conservative 25; Mismatches 63; Indels 35; Gaps 10;

QY 4 EYSELKPRVYVTONGKEMSTIVSEDFILPVYKGELEKGYOPDMEISG--FEKK 60  
DB 42 KTQETKED-----ENKSNNNVNSSENF-MKIYKNLIKQYHVEKKQKQENIESEK 95  
QY 61 DA-----GVYINLSKDTPIKPVFKKIEEKEENKPTFDVSKKQDN--POV--NH--S 107  
DB 96 KATSKKKIKKLNFKLKKNNLLEKTVK--KAFELAKTVFSPSLKKYVTPITSNNHIS 152  
QY 108 QLNESHKED-----LQREHSQKSDSTKYDTATVLDKNNISSKSTNNPN 153  
DB 153 QNDHKKKKQDKKIKKIKLIEKSNKDDNN-----NINNKNNQKKCDTNNIN 198

RESULT 33  
09OV00  
ID 09OV00 PRELIMINARY; PRT; 157 AA.  
AC 09OV00;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE CG31958 protein.  
GN CG31958 OR CG10022.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydriidae; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt J.C., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borzova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merklow G., Mishina N.V., Mobarry C., Morris J., Mostreli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorset V., Doup L.B., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Mostreli A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Pargass V., Park S., Patel S., Pfeiffer B.,  
 RA Phouenavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.T., Bergman C., Berman B., Carlson J.W., Celinker S.E.,  
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Seearle S.M., Smith E., Shu S., Smutnak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AAC03578; AAF51062.2; -;  
 DR HSSP; P02593; 1CTR.  
 DR FLYbase; FBgn0051958; CG31958.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00036; efhand; 3;  
 DR ProDom; PD000012; EF-hand; 1.  
 DR SMART; SM00054; EFh; 3.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 KW Calcium; Calcium-binding.

SQ SEQUENCE 157 AA; 18263 MW; 543F0480E11D9EC1 CRC64;  
 Query Match 12.3%; Score 98; DB 5; Length 157;  
 Best Local Similarity 22.7%; Pred. No. 3.8;  
 Matches 34; Conservative 29; Mismatches 53; Indels 34; Gaps 5;  
 QY 23 MSSTVSEEDFLIPYKGLKGYQFDGMEISGFEKKDAGYI-----NLSKDTF 73  
 DB 1 MELSVSEDDILKNIT-SLIDK-----DNEGATSKELGVIALGRQPNESIAREF 52  
 QY 74 IPEVFKIIE-KKEENKTFPVSKKD-----PQVHSLNESHRE 116  
 DB 53 CNVIIRKMDTKREBELRDAFVPEPKENGYISTELRAVFNALGEKDELDELMIRY 112  
 QY 117 DIQREHSQKSDTKDVTATVLDKNNISK 146  
 DB 113 DLDQDHNIFESNNMTPIINVTAVK 142  
 RESULT 34  
 ID 0813J9 PRELIMINARY; PRT; 4524 AA.  
 AC 0813J9;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PFI325W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_Taxid:36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrett B.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2255708; PubMed=12368867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphray S., Jagels K., James K.D., Johnson D., Kerhrou A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Leonard N.,  
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinovitch E.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Yalston J.E., Craig A., Newbold C., Barrett B.G.;  
 RT "Sequence of plasmodium falciparum chromosomes 1, 3-9 and 13.";  
 RL Nature 419:527-531(2002).  
 DR EMBL; AL929354; CAD51629.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 4524 AA; 525287 MW; 542206173C29567A CRC64;  
 Query Match 12.3%; Score 98; DB 5; Length 4524;  
 Best Local Similarity 24.8%; Pred. No. 1.5e+02;  
 Matches 40; Conservative 26; Mismatches 75; Indels 20; Gaps 6;  
 QY 2 TGEVSELPKRVTVTI-----ONGKMSSTVSEDFLIPYKGLKGYQFDGMEI 53  
 DB 3832 TGEVSELPKRVTVTI-----ONGKMSSTVSEDFLIPYKGLKGYQFDGMEI 3888  
 QY 54 SGFEKQAGYVINSKDTFIKPVKKEEENKTFPVSKKDQPVQVHSLNESH 113  
 DB 3889 SNFNISLIDN-MINLNKN-SITTTDKIKSCCHNRMSISIKKIRSVYIKIS 3946  
 QY 114 RKEDLQREHSQKSDTKDVTATVLDKNNISKSTTNPNK 154

DB 3947 MNIIQKET-EKKQDTYKEV-----KNLKKKSKDANKKNK 3980

## RESULT 35

081HW3 PRELIMINARY; PRT; 449 AA.

AC 081HW3; MEDLINE=22255705; PubMed=12368864;

DT 01-MAR-2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DE Hypothetical protein.

GN PF11\_0413.

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OK NCBI\_TaxId=36329;

RP SEQUENCE FROM N.A.

RC STRAIN=3D7;

RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,

RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,

RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Xys S.,

RA Chan M.-S., Nene V., Shalim S.J., Sub B., Peterson J., Angiuoli S.,

RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,

RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,

RA Venner J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

RA Fraser C.M., Barrel B.

RT "Genome sequence of the human malaria parasite Plasmodium

RT falciparum."

RL Nature 419:498-511 (2002).

DR EMBL; AB014842; AAN3596.1; --

KW Hypothetical protein.

SO SEQUENCE 449 AA; 52812 MW; 85B62272D6257C68 CRC64;

Query Match 12.2%; Score 97.5; DB 5; Length 449;

Best Local Similarity 25.1%; Pred. No. 13;

Matches 42; Conservative 28; Mismatches 48; Indels 49; Gaps 7;

DB 21 KEMSTTIVSEDFILPVY-----KGELEKGYOPDGMISGFEKKDAGVYINLSKDTF 73

DB 55 KELTITILKEK--ILIVNEQNEKKKKKKKQFQSD---ISKKKSKXKLDKSSSKNN 109

OY 74 IKPVEFKIE-----EKKEE-----NKPTFDV-----SKKKNPQVN 105

DB 110 LKKKKKIESEBETINSDDEEEBEGYKKQKKQKNSVSTLSLEKKKKKKRSESSN 169

OY 106 HSQLNEHKKEDLOREHSHQKSDSTKDVATVLDKNNISSKSTNNP 152

DB 170 NNDVNEEYDEDDDEQEEEEEES-----LSKKSKKKKSTTSP 207

## RESULT 36

095S93 PRELIMINARY; PRT; 556 AA.

AC 095S93; MEDLINE=20512582; PubMed=11058132;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE GM052229P.

GN ASPH OR CG8421 OR CG18658.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OK NCBI\_TaxId=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nunoo J., Pacleb J., Paragas V., Park S., Pionanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;

RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL; AY060905; AAL28453.1; --

DR Flybase; FBgn0034075; Asph.

SO SEQUENCE 556 AA; 63089 MW; 95D82EAC57D11FE8 CRC64;

Query Match 12.2%; Score 97.5; DB 5; Length 556;

Best Local Similarity 24.5%; Pred. No. 17;

Matches 34; Conservative 28; Mismatches 52; Indels 25; Gaps 5;

OY 31 EDFILPVYKGELEKGYOPDGW-----EISGFEKKDAGVYI-----NLSKDTIK 75

DB 78 EDLDTPLSESRFSK--VPDGMVDEHDEHDHGDVQEPGEALDHDHDDHDEDEEE 135

OY 76 PVFKIEEKKEEENKPT-----FVSKKNPQVNHSQLNESHKEDLOREHSHQKSDS 129

DB 136 PLTELEEELEEEEPTEDEPPADEEYEEDEENNA--GENITABDAVEEEEDND 193

OY 130 TKDVATVLDKNNISSKST 148

DB 194 EGTVAIVATTEATTEAT 212

## RESULT 37

09K5S1 PRELIMINARY; PRT; 614 AA.

AC 09K5S1; MEDLINE=20512582; PubMed=11058132;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Hypothetical protein BH4017.

GN BH4017.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OK NCBI\_TaxId=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fujii F., Hirama C., Nakamura Y., Ogasaawara N., Kuhara S.,

RA Horikoshi K.; "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis."

RL Nucleic Acids Res. 28:4317-4331 (2000).

DR EMBL; AP001520; BAB07736.1; --

DR InterPro; IPR006829; Transposase.30.

DR Pfam; PF04740; Transposase.30; 1-

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 614 AA; 70704 MW; 2E3FDEEC74E0294C CRC64;

Query Match 12.2%; Score 97.5; DB 16; Length 614;

Best Local Similarity 27.3%; Pred. No. 18;

Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;

OY 6 SELKPRVTVIIONKEMSTTIVSE-----EDFILPVYKGELEKGYOPDGW-----EISG 55

DB 386 SEPKPE--TYLTQTAIMQ-TPIVNEYSPTREBEP-----ARKAQLDGMADVSKVY 435

OY 56 FEGKKDAGVYINLSKDTIKPVFKIEEKK--EEENKPTFDVSKK---DNPQVNHSQL 109

DB 436 FAGRNIAQL-----GKIEEKKQDKNTNNTTFDFPKKEVNVVOSPIKSTAL 482

OY 110 NESHKEDLOREHSHQKSDSTKDVATVLD 139

DB 483 GKIVIGIVDLGRKHKXKEDLRRRLSKSGL 512

## RESULT 38

093424 PRELIMINARY; PRT; 385 AA.

AC 093424; MEDLINE=20512582; PubMed=11058132;

DT 01-JAN-1999 (TREMBLrel. 09, Created)

DT 01-JAN-1999 (TrEMBLrel. 09, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE Hypothetical glycine-rich 37.0 kDa protein E02A10.2 in chromosome V precursor.  
 GN E02A10.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NC NCBL\_Taxid=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Thomas K.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN=BRISTOL N2;  
 RA Jones S.J.M.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: 281053; CAB02877.1; -.  
 DR HSSP; P10968; 2CWG; CE09116.  
 DR WormPep; E02A10.2; EGShell.  
 DR InterPro; IPR002952; EGShell.  
 DR PRINTS; PR01228; EGShell.  
 KM Hypothetical protein; Signal; Transmembrane.  
 FT SIGNAL 1 17  
 FT CHAIN 18 385  
 FT TRANSMEM 165 185  
 FT DOMAIN 23 190  
 FT SEQUENCE 385 AA; 36963 MW, 32AB25AE9B5073FE CRC64;  
 SQ  
 Query Match 12.1%; Score 97; DB 5; Length 385;  
 Best Local Similarity 29.6%; Pred. No. 12;  
 Matches 42; Conservative 24; Mismatches 48; Indels 28; Gaps 7;  
 QY 1 DNGEVELKPHRYTVIQ--NGKMSSTIVSEDFILPYKGLKGYQFDGWEISGFSG 58  
 DB 256 ETDDFPSLQENVHCTILINHKVARK--NEEDKKEEKKEEKEEVEKKE--EDB 310  
 QY 59 KKDAGVINLSKDTFIPVKFKLIEKKKEENKPTFDVSKKKNPQVNSQLNESHKEDL 118  
 DB 311 KKDE-----EP--KKEEKKKEEKK--EVEKKE-----EKKDEPKKEE 349  
 QY 119 QREHSQKSDSTKDTATVLDK 140  
 DB 350 KKEEKEKDEVEKESEKVEEK 371  
 RESULT 39  
 ID Q815T1 PRELIMINARY; PRT; 558 AA.  
 AC Q815T1;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE Hypothetical protein.  
 GN PFL0600W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NC NCBL\_Taxid=36329;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RA MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
 RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Valdivya A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carnucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

RA Fraser C.M., Barrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium falciparum."  
 RL Nature 419:498-511 (2002).  
 DR EMBL; AE014846; AAN36209.1; -.  
 KM Hypothetical protein.  
 SQ SEQUENCE 558 AA; 66487 MW; 2AD029EA69F90D3 CRC64;  
 Query Match 12.1%; Score 97; DB 5; Length 558;  
 Best Local Similarity 22.4%; Pred. No. 18;  
 Matches 35; Conservative 31; Mismatches 64; Indels 26; Gaps 4;  
 QY 18 QNGKMSSTIVSEDFILPYKGLKGYQFDGWEISGFSGKDGAGYVINSKDTFIRV 77  
 DB 41 QNGKTYTYNNLSKKNNNGIIXGSKHNGHLF-----INNYKKRNVKYKLNKYPGSGF 96  
 QY 78 FKIEKKKEENKPTFDVSKKKNPQVNSQLN-----SHRKEKL 118  
 DB 97 FKK-KDKNDKNDK--DKTHMKDNGNLNKKKEGKIGKQKEKEKEKILIKDKKQKDE 153  
 QY 119 QREHSQKSDSTKDTATVLDKNNISSKSTTNPK 154  
 DB 154 KKKKELQKEKKTKLTNLVKNNKNNKNNKNNK 189  
 RESULT 40  
 ID Q9STN4 PRELIMINARY; PRT; 988 AA.  
 AC Q9STN4;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)  
 DE Hypothetical 109.0 kDa protein.  
 GN T8B5.30 OR AT4G08340.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NC NCBL\_Taxid=3702;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,  
 RA Barrell B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K.,  
 RA Schellier C.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL109819; CAB52556.1; -.  
 DR EMBL; AL161511; CAB77959.1; -.  
 DR InterPro; IPR001760; Opsin.  
 DR InterPro; IPR003653; SMOO protease.  
 DR Pfam; PF02302; Peptidase\_C46; 1.  
 DR PROSITE; PS00238; OPSIN; 1.  
 DR PROSITE; PS06000; ULP\_PROTEASE; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 988 AA; 108955 MW; 01C518587D460EAD CRC64;  
 Query Match 12.1%; Score 97; DB 10; Length 988;  
 Best Local Similarity 22.1%; Pred. No. 34;  
 Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;  
 QY 19 NQKMSSTIVSEDFILPYKGLKGY-----QFDGWEISGFSGKDGAGYVINSKDTF 73



Db 383 NGRQNSNVQSSVDIELSYTTPKVPBGVLNVSERDIVELEVVEDVRSAGGLSPNVQORDN- 441  
Qy 74 IKPVFKIEEKE-----EENKPTFDVSKKQNPQVNHSQLNESHREDLQREE 122  
Db 442 VEPVGGDDVRSAGDMSPNPSAANNVREGPATFDIMESEDNPGRDVNAFPMEDHIRSEVQLSP 501  
Qy 123 HSQKSDSTKDYT--ATVLDKNNISSKSTNNPNK 154  
Db 502 HVL---GAKDYTDVSDPTDKVGVDVTDASDPT 532

Search completed: February 10, 2004, 10:57:07  
Job time : 28.2661 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 37.3096 Seconds  
(without alignments)  
697.707 Million cell updates/sec

```

Title: US-10-067-385-8_COPY_610_773
Perfect score: 848
Sequence: 1 TTVKEFILNKDTGEVSELPK.....ATVLDKNNISSKSTNNPNK 164

```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched:      1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

A\_Geneset\_19Jun03:\*

1:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1980.DAT:*
2:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1981.DAT:*
3:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1982.DAT:*
4:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1983.DAT:*
5:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1984.DAT:*
6:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1985.DAT:*
7:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1986.DAT:*
8:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1987.DAT:*
9:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1988.DAT:*
10:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1989.DAT:*
11:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1990.DAT:*
12:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1991.DAT:*
13:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1992.DAT:*
14:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1993.DAT:*
15:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1994.DAT:*
16:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1995.DAT:*
17:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1996.DAT:*
18:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1997.DAT:*
19:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1998.DAT:*
20:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA1999.DAT:*
21:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA2000.DAT:*
22:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA2001.DAT:*
23:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA2002.DAT:*
24:	/SIDS1/gcgdata/genseq/genseqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	848	100.0	773	22	AA848343	S. pneumoniae Spi3
2	848	100.0	2120	21	AA818110	Streptococcus pneu
3	848	100.0	2140	24	ABU01020	S. pneumoniae type
4	615	72.5	117	19	AAW55096	Streptococcus pneu
5	615	72.5	117	23	ABP54590	S. pneumoniae Spi04
6	119	14.0	746	22	AA811779	S. epidermidis ope
7	119	14.0	778	23	ABP39023	Streptococcus epi
8	107.5	12.7	484	21	AA047777	Arabidopsis thalia
9	106	12.5	665	21	AA818278	Plasmodium falcipa

10	101.5	12.0	564	22	AB8615917
11	99.5	11.7	1384	24	ABP554133
12	99.5	11.7	1404	24	ABP553933
13	97.5	11.5	2519	22	ABG166366
14	96.5	11.4	408	21	AAAG371344
15	96.5	11.4	456	21	AAAG371133
16	96.5	11.4	476	21	AAAG371232
17	96.5	11.4	2060	23	AAE30967
18	96	11.3	2500	21	AAE182722
19	93.5	11.0	645	24	ABJ191066
20	93.5	11.0	654	24	ABP56879
21	92.5	10.9	645	24	ABJ189799
22	92	10.8	239	22	AAAG23977
23	92	10.8	309	23	ABP408822
24	91	10.7	442	23	ABP281868
25	91	10.7	902	23	ABP40312
26	90.5	10.7	209	21	AAE428977
27	90.5	10.7	635	23	ABP732099
28	90.5	10.7	757	21	AAAY43644
29	90.5	10.7	788	21	AAE531319
30	90.5	10.7	801	23	ABP589633
31	90.5	10.7	922	22	AAAG50233
32	90	10.6	1183	22	ABE587656
33	89.5	10.6	402	20	AAE318750
34	89.5	10.6	511	20	AAAY35091
35	89.5	10.6	516	23	ABP735707
36	89.5	10.6	525	20	AAAY03189
37	89.5	10.6	1408	22	ABE587044
38	89	10.5	296	23	ABE933411
39	88.5	10.4	281	22	ABE648288
40	88.5	10.4	1141	22	AAAG50088
41	88	10.4	225	23	ABP739922
42	88	10.4	258	22	ABE458444
43	88	10.4	817	22	AAAY79318
44	88	10.4	817	22	AAAY79319
45	88	10.4	2515	22	ABE128218

## ALIGNMENTS

## RESULT 1

ID AAB48343 standard; Protein; 773 AA.

AC AAB48343;

DT 20-APR-2001 (first entry)

DE S. pneumoniae Sp130 polypeptide

KW Immunogenic; Sp128; pneumococcal; otitis media; nasopharyngeal;

KW antibacterial; auditory; vaccine.

05 Streptococcus pneumoniae.

PN WO200076540-A2.

PD 21-DEC-2000

PF 09-JUN-2000

PR 10-JUN-1999; 99US-0138453

PA (MEDI-) MED IMMUNE INC.

PI Adamou JE, Choi GH;

DR WPI; 2001-112197/12.

XX

Drosophila melanogaster  
Human MDR1-22 protein  
Human MDR1-22 protein  
Novel human diatom  
Arabidopsis thaliana  
Arabidopsis thaliana  
Arabidopsis thaliana  
Staphylococcus aureus  
Plasmodium falciparum  
Pathogen specific  
Staphylococcus epidermidis  
Pathogen specific  
S. epidermidis open  
Staphylococcus epidermidis  
Staphylococcus epidermidis  
Scaphylococcus epidermidis  
Human ORF ORF2661  
Candida albicans  
Human cell cycle 1  
Human colon cancer  
Human polypeptide  
Shrimp white spot  
Drosophila melanogaster  
S. aureus trigger  
Chlamydia pneumoniae  
Candida albicans  
S. aureus trigger  
Drosophila melanogaster  
Herbicidally active  
Drosophila melanogaster  
Shrimp white spot  
Candida albicans  
Human protein sequence  
Human protein SEQ  
Human protein SEQ  
Human secreted protein

PT preventing pneumococcal infections, particularly infections caused by  
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
PT blood infections -

PS Claim 8; Page 51-54; 54pp; English.

XX The invention relates to novel immunogenic polypeptides, Sp128 and Sp130  
XX from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
CC the treatment and prevention of pneumococcal infections, particularly  
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
CC bronchial, lung or blood infections. The antigens are used as immunogenic  
CC agents to stimulate an immune response. The antisera and antibodies may  
CC also be used in diagnosing and treating pneumococcal infections.  
CC Recombinant polypeptides serve as a mechanism for stimulating production  
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Sp130 polypeptide.

SQ Sequence 773 AA;

Query Match 100.0%; Score 848; DB 22; Length 773;

Best Local Similarity 100.0%; Pred. No. 1.1e-74;

Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEFLNKDNGESELKPHRYVTITONGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60

DB 610 TTVKEFLNKDNGESELKPHRYVTITONGKMSSTIVSEEDFILPVYKGELEKGYQFDG 669

QY 61 WEISGEGKKDAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSKKKDNQVNHSQLN 120

DB 670 WEISGEGKKDAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSKKKDNQVNHSQLN 729

QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164

DB 730 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 773

RESULT 2

AA81710

ID AAY81710 standard; Protein; 2120 AA.

AC AAY81710;

DT 02-JUN-2000 (first entry)

DE Streptococcus pneumoniae protein sequence ID3.

XX Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;

KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;

KW kidney disease; diabetes; immunosuppressive disorder; otitis media;

XX pneumococcal septicemia; sinusitis; meningitis; therapy.

OS Streptococcus pneumoniae.

PN WO200006738-A2.

PD 10-FEB-2000.

PF 27-JUL-1999; 99WO-GB02452.

PR 27-JUL-1998; 98GB-0016336.

PA 19-MAR-1999; 99US-0125329.

PI (MICR-) MICROBIAL TECHNIQS LTD.

PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;

XX WPI: 2000-195301/17.

DR N-PSDB; AA291806.

PT Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections -  
XX

PS Claim 2; Page 41-42; 76pp; English.

XX This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicemia, otitis media, sinusitis, and  
CC meningitis.

SQ Sequence 2120 AA;

Query Match 100.0%; Score 848; DB 21; Length 2120;

Best Local Similarity 100.0%; Pred. No. 4.3e-74;

Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEFLNKDNGESELKPHRYVTITONGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60

DB 1923 TTVKEFLNKDNGESELKPHRYVTITONGKMSSTIVSEEDFILPVYKGELEKGYQFDG 1982

QY 61 WEISGEGKKDAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSKKKDNQVNHSQLN 120

DB 1983 WEISGEGKKDAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSKKKDNQVNHSQLN 2042

QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164

DB 2043 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2086

RESULT 3

ABU01020

ID ABU01020 standard; Protein; 2140 AA.

AC ABU01020;

DT 11-FEB-2003 (first entry)

DE S. pneumoniae type 4 strain protein from coding region #590.

XX Bacterial meningitis; pneumonia; sepsis; otitis media;

KW ear infection; antiinflammatory; antibacterial; immunostimulant;

KW auditory; respiratory; gene therapy; vaccine.

XX Streptococcus pneumoniae type 4 strain.

PN WO200277021-A2.

PD 03-OCT-2002.

PF 27-MAR-2002; 2002WO-IB02163.

PR 27-MAR-2001; 2001GB-0007658.

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Masignani V, Tettelin H, Fraser C;

XX WPI: 2003-040579/03.

DR N-PSDB; ABX06302.

PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media

PT or ear infection -

PS Claim 1; SEQ ID No 1180; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%  
PS identity to any of the 2469 amino acid sequences, identified in the  
XX specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the  
CC proteins, treating a patient by administering the protein, DNA or  
CC antibody (in a composition), a kit comprising first and second primers,  
CC which are the nucleic acid cited above or fragments between nucleotides  
CC 8-100 of a sequence not defined in the specification, for amplifying a  
CC target sequence contained within a Streptococcus nucleic acid sequence,  
CC where the first primer is substantially complementary to the target  
CC sequence and the second primer is substantially complementary to the  
CC complement of the target sequence, and where the parts of the primers  
CC having substantial complementarity define the termini of the target  
CC sequence to be amplified, assay comprising contacting a test compound  
CC with the protein, and determining whether the test compound binds to the  
CC protein and a Streptococcus pneumoniae bacterium, where one or more  
CC genes encoding the proteins has been rendered inactive. The proteins,  
CC nucleic acid molecules, antibody and compositions are useful as  
CC medicaments for treating or preventing a disease or infection due to  
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
CC sepsis, otitis media or ear infection. They are also useful in developing  
CC vaccines, diagnostics and antibiotics. The methods are useful for  
CC identifying immunodominant proteins. The present sequence is one of  
CC the 2469 proteins expressed by the identified coding regions from the  
CC genomic sequence.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SO Sequence 2140 AA;

Query Match 100.0%; Score 848; DB 24; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 4,4e-74;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTVEFIANKDTGSEVSELPKPRVVTYTIQNGKMSSTIVSEDFILPVYKGELEKGYODG 60

DB 1943 TTVEFIANKDTGSEVSELPKPRVVTYTIQNGKMSSTIVSEDFILPVYKGELEKGYODG 2002

OY 61 WEISGFEKKGAGYVYINISKDTFIKPVFKIIEKKEENKPTFDVSKKKDNPOVNHSQLN 120

DB 2003 WEISGFEKKGAGYVYINISKDTFIKPVFKIIEKKEENKPTFDVSKKKDNPOVNHSQLN 2062

OY 121 ESHRKEDLQREHSHQKSDSTKDVTATVLDKNNISSKSTNNPNK 164

DB 2063 ESHRKEDLQREHSHQKSDSTKDVTATVLDKNNISSKSTNNPNK 2106

RESULT 4

AAM55096 standard; Protein; 117 AA.

AC AAM55096;

DT 02-OCT-1998 (first entry)

DE Streptococcus pneumoniae SP0043 protein.

KM Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
KW detection; pneumonia; otitis media; meningitis.

OS Streptococcus pneumoniae.

PN MO9818930-A2.

PD 07-MAY-1998.

XX 30-OCT-1997; 97MO-US19422.

XX 31-OCT-1996; 96US-0029960.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

DR WPI, 1998-272224/24.

DR N-PSDB; AAV27357.

PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
PT pneumoniae - or their epitope-containing fragments, useful in  
PT protective or therapeutic vaccines, and for diagnosis

PS Claim 11; Page 62; 118pp; English.

CC The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.

SO Sequence 117 AA;

Query Match 72.5%; Score 615; DB 19; Length 117;  
Best Local Similarity 100.0%; Pred. No. 7.3e-53;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 YKGELEKGYODGMEISGFEKKGAGYVYINISKDTFIKPVFKIIEKKEENKPTFDVSK 107

DB 1 YKGELEKGYODGMEISGFEKKGAGYVYINISKDTFIKPVFKIIEKKEENKPTFDVSK 60

OY 108 KKNPQVNHSQLNESHKREDLQREHSHQKSDSTKDVTATVLDKNNISSKSTNNPNK 164

DB 61 KKNPQVNHSQLNESHKREDLQREHSHQKSDSTKDVTATVLDKNNISSKSTNNPNK 117

RESULT 5

ABP54590 standard; Protein; 117 AA.

AC ABP54590;

DT 04-SEP-2002 (first entry)

DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.

KM Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
KW antibacterial; Streptococcal infection; detection.

OS Streptococcus pneumoniae.

PN US2002061545-A1.

PD 23-MAY-2002.

PF 22-JAN-2001; 2001US-0765272.

PR 30-OCT-1997; 97US-0961083.

PA (CHOI/) CHOI G H.

PA (KUNSCH/) KUNSCH C A.

PA (BARA/) BARASH S C.  
PA (DILL/) DILLON P J.  
PA (DOUG/) DOUGHERTY B.  
PA (FANN/) FANNON M R.  
PA (ROSE/) ROSEN C A.  
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,  
PI Rosen CA;  
XX WPI, 2002-479261/51.  
DR N-PSDB; AB084825.  
PT New Streptococcus pneumoniae antigens, useful for detecting  
PT Streptococcus and for preventing or attenuating disease caused by  
PT Streptococcus infection -  
XX Claim 11; Page 29; 70pp; English.  
XX  
XX AB084792 to AB084904 represents nucleic acids which encode the  
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
CC The S. pneumoniae antigens have antibacterial activity and can be  
CC used in vaccines. The S. pneumoniae antigens can also be used to  
CC prevent or attenuate a Streptococcal infection in an animal. The  
CC polynucleotides encoding the S. pneumoniae antigens can be used to  
CC detect Streptococcus nucleic acids. AB084905 to AB085130 represent  
CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
CC which are used in an example from the present invention.  
XX  
SQ Sequence 117 AA;  
Query Match 72.5%; Score 615; DB 23; Length 117;  
Best Local Similarity 100.0%; Pred. No. 7.3e-53;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 YKGELEKGYQFDGMEISGFEKKDAGYVNLKDTFIKPVFKKIEKKEENKPTDVS 107  
DB 1 YKGELEKGYQFDGMEISGFEKKDAGYVNLKDTFIKPVFKKIEKKEENKPTDVS 60  
QY 108 KKDNPQVNHSQLNESHKEDLOREHESOKSDTKVATVLDKNNISSKSTNN 164  
DB 61 KKDNPQVNHSQLNESHKEDLOREHESOKSDTKVATVLDKNNISSKSTNNPK 117  
RESULT 6  
AA081779  
ID AA081779 standard; Protein; 746 AA.  
XX  
AC AA081779;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.  
XX  
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;  
KW vaccination; endocarditis.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN WO200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000WO-US30782.  
XX  
PR 09-NOV-1999; 99US-0164258.  
XX  
PA (GLAX) GLAXO GROUP LTD.  
XX  
PI kimmerly WJ;  
XX  
DR WPI; 2001-316495/33.  
DR N-PSDB; AAH52629.  
XX

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX  
XX Claim 18; Page 208; 218pp; English.  
XX  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AA081454 to AA083120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX  
SQ Sequence 746 AA;  
Query Match 14.0%; Score 119; DB 22; Length 746;  
Best Local Similarity 27.0%; Pred. No. 0.0081;  
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
QY 10 KDTGEVSEIKPRVVTITONGKEMSTIYSEEDFILPVK-----GELEKGYQFDG 61  
DB 585 EDSVANAQLK-----ITIGNGKQIKQGSVSGTKVLPKSKVLMWTDGELTMP-DMTGWTK 639  
QY 62 -EISGFE-----GKQDAGYVNLKDTFIKPVFKKIEKKEENKPTDVS---K 107  
DB 640 EYVLAPELITKLVKSTKNGFVTNOSISKQIYK-----NKDKLEVISLSAED 686  
QY 108 KKDNPQVNHSQLNESHKEDLOREHESOKSDTKVATVLDKNNISSKSTNN 161  
DB 687 TDDQEKTEDESDNKSXKDKADEHDSNTSSSTKX-----DKSNADSKNDSD 734  
RESULT 7  
ABP39023  
ID ABP39023 standard; Protein; 778 AA.  
XX  
AC ABP39023;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.  
XX  
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX

DR WPI: 2002-381255/41.  
DR N-PSDB; ABN91568.  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 3868; 267pp; English.  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP3124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
CC  
SQ Sequence 778 AA;  
Query Match 14.0%; Score 119; DB 23; Length 778;  
Best Local Similarity 27.0%; Pred. No. 0.0086;  
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
Qy 10 KDTGEVSEIKPHRTVTITQNGKMSSTIYSEDFILPYVK-----GELEKGYQPDGW-- 61  
Db 617 EDSVNAQSLKP-----ITIGNGKQIKQGSYKSGTKVLPFSKVMMLMTDGLTWP-DMTGWTK 671  
Qy 62 -EISGFE-----GKDAQGVYIN--LSMDTFIKVFKEKKEENKFTDVS---K 107  
Db 672 EDVLAPELDTTKIKVSTKNGFVTNOSISKQILK-----NDKTEVSLASAE 718  
Qy 108 KKDQNVNHSQLESRKEDLQREHSQKSDSTKDTATVLDKNNISKSTTN 161  
Db 719 TDDDEKTEDESDSKSKKDKADEHSHNNTSSSTK-----DKSNADSNDD 766  
RESULT 8  
AAG47777  
ID AAG47777 standard; Protein; 484 AA.  
XX  
AC AAG47777;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60255.  
XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
FN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PE 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134220.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.





PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
 PT diagnosis of P.falciparum infection -  
 XX Disclosure; Page 321-322; 577pp; English.  
 XX  
 CC The present invention describes proteins and their fragments (I) encoded  
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
 CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)  
 CC vaccines against P. falciparum infection comprising (i) or (ii).  
 CC (i) and (ii) are useful for the development of vaccines against  
 CC P. falciparum infection. (i) and polyclonal antisera or a monoclonal  
 CC antibody raised to immunogens comprising the sequences of (i), are  
 CC useful in the detection of infection with P. falciparum. Furthermore,  
 CC (i) (especially when they are rifins or secreted or membrane proteins)  
 CC can aid the identification of drugs to treat or prevent P. falciparum  
 CC infection, or they can be used to identify drug resistance in  
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
 CC subsequent identification of proteins encoded by it will help to expand  
 CC our understanding of parasite biology, a process hampered by the  
 CC complexity of the parasitic lifecycle, and provide new targets for  
 CC vaccine and drug development. Parasite resistance to drugs and mosquito  
 CC resistance to insecticides have led to a resurgence of malaria in many  
 CC parts of the world, and there is a pressing need for vaccines and new  
 CC drugs. AA470078 to AA470287 and AAB18144 to AAB18352 represent nucleotide  
 CC and protein sequences given in the present invention, but which are not  
 CC specifically mentioned within the specification.  
 CC  
 XX Sequence 665 AA;  
 SQ  
 Query Match 12.5%; Score 106; DB 21; Length 665;  
 Best Local Similarity 24.3%; Pred. No. 0.13; Indels 56; Gaps 10;  
 Matches 42; Conservative 35; Mismatches 40;  
 QY 10 KDTGEVSEIKPRVAV-VTIQNGKEMSVTVSEDEPILPYKGLERKGYQPDGMEI--SGF 66  
 DB 127 EEKNNKINSDDLHROEMLNQSGK-----NEQDI-----NKNKKGQ---DLSNSVA 169  
 QY 67 ECKKQAGVYINISKOTFIKVPFKIEKKE-----EENKPTPD---VSKKQKNP 112  
 DB 170 EENKXQ-----VKEGVKELEKKEKKEKISDHHKVENKKSDDHKVVENKKSDDH 217  
 QY 113 QVNHSQLNESHREKEDLQR-EESHQKSDSTKDTATVLDKNNTSSKSTNNPK 164  
 DB 218 KVENKKSDDHKIEVKVEHEDEDE-----DKKEKSSKNKNENK 261  
 RESULT 10  
 ABB61977  
 ID ABB61977 standard; Protein; 564 AA.  
 XX  
 AC ABB61977;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 12723.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;

XX  
 DR WPI; 2001-6556860/75.  
 DR N-P8DB; ABL06080.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 CC Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.  
 XX  
 PS The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB161511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 564 AA;  
 SQ  
 Query Match 12.0%; Score 101.5; DB 22; Length 564;  
 Best Local Similarity 24.5%; Pred. No. 0.29; Indels 25; Gaps 5;  
 Matches 34; Conservative 29; Mismatches 51;  
 QY 41 EDFILPVYKGLERKGYQPDGMEI-----EISGFEKQKQGYI-----NLSKPTPK 85  
 DB 78 EDLDTPLSESRFSK--VFDGVDNDRHDDHDDVQEPGEGLDDHDEHDDHDEDEDE 135  
 QY 86 PVFKKIEEKKEENKPT-----PDVSKKQNPQVNHSQLNESHREKEDLQREHSHQKSDS 139  
 DB 136 PLTELELELELEBEPEDEDEPADDEYEDEDEBNNA--GENTIADEEEEBEEDMD 193  
 QY 140 TKDVTATVLDKNNTSSKST 158  
 DB 194 EGTVAIVTAETEAETEAET 212  
 RESULT 11  
 ABB55413  
 ID ABB55413 standard; Protein; 1384 AA.  
 XX  
 AC ABB55413;  
 XX  
 DT 04-FEB-2003 (first entry)  
 XX  
 DE Human MDDT-22 protein SEQ ID NO:22.  
 XX  
 KW Human; MDDT; molecules for disease detection and treatment; anti-HIV;  
 KW antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nocotropic;  
 KW anticonvulsant; antinfertility; antianorectic; antisthmatic;  
 KW immunosuppressive; antithyroid; cytotatic; hepatotropic; dermatological;  
 KW antidiabetic; nephrotoxic; antipain; thyromimetic; neuroprotective;  
 KW osteopathic; antiarthritic; antiparastitic; antihelminthic; antiprositic;  
 KW uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;  
 KW virucide; protozoacide; fungicide; gene therapy; cell proliferative;  
 KW cancer; developmental disorder; neurological disorder; infection;  
 KW reproductive disorder; autoimmune disorder; inflammatory disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200278420-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 29-MAR-2002; 2002WO-US09809.  
 XX  
 PR 30-MAR-2001; 2001US-280387P.  
 XX  
 PR 05-APR-2001; 2001US-282335P.  
 XX  
 PR 13-APR-2001; 2001US-283663P.



Best Local Similarity 21.7%; Pred. No. 1.6;  
Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;  
QY 4 KEFINKDGEVSELKPHVVTWVTIONGKMSSTIVSEEDFILPYKGELEKGYOPDMEI 58  
DB 1033 KEFIMQWQAEIKKLSAEVSLKEAYHMEMISHQENHAKWIS--ADSQKSSVQQLNEQLKKA 1091  
QY 59 -----DQWEISGPEGKKDA-----GVVNLMSK----- 80  
DB 1092 QVONSHTELAARHQOVQAGREIERLSSELDKQKSKDAGHNLAEIGASKVREAH 1151  
QY 81 -----DTFIKPVFKKIEKKK-----EENKPTFDVSKKKDNPQVNSQLNESHRE 126  
DB 1152 LEARMQAEIKKLSAEVSLKEAYHMEMISHQENHAKWIS--ADSQKSSVQQLNEQLKKA 1209  
QY 127 DQREBHSQKSDSTKDVATVYLDKNI 153  
DB 1210 KLEEBE--AQDTVSNLHQOVQDRENV 1233  
RESULT 13  
ABG16636  
ID ABG16636 standard; Protein; 2519 AA.  
AC  
XX ABG16636;  
XX  
XX 18-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #16627.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HSE-) HSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX MPI: 2001-639362/73.  
XX DR N-PSDB; AAS80823.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity  
XX  
XX Claim 20; SEQ ID NO 46995; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPD  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 2519 AA;  
SQ  
Query Match 11.5%; Score 97.5; DB 22; Length 2519;  
Best Local Similarity 24.8%; Pred. No. 5.8;  
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;  
QY 4 KEFINKDGEVSELKPHVVTWVTIONGKMSSTIVSEEDFILPYKGELEKGYOPDMEI 63  
DB 635 EKVWYKDKKPVVTEKTPSVTEKEVSKKEPS-----PV-KAEVA-----EK 674  
QY 64 SGFEGKDAQVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKDNPQVNSQLNESH 123  
DB 675 QATDVKPPAAKKEKTVKKEKTKVP-----EDKKEKEKPKKVAKKEDKTPI---KKEK 726  
QY 124 RKEDLQRE-----EHSQKSDSTKDV 143  
DB 727 KKEVYKKEVYKKEIKKEKKEPKKEV 751  
RESULT 14  
AAG37134  
ID AAG37134 standard; Protein; 408 AA.  
XX  
XX AAG37134;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 45610.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 25-MAR-1999; 99US-0126264.  
XX PR 29-MAR-1999; 99US-0126785.  
XX PR 01-APR-1999; 99US-0127462.  
XX PR 06-APR-1999; 99US-0128234.  
XX PR 08-APR-1999; 99US-0128714.  
XX PR 16-APR-1999; 99US-0129845.  
XX PR 19-APR-1999; 99US-0130077.  
XX PR 21-APR-1999; 99US-0130510.  
XX PR 23-APR-1999; 99US-0130891.  
XX PR 28-APR-1999; 99US-0131449.  
XX PR 30-APR-1999; 99US-0132048.  
XX PR 30-APR-1999; 99US-0132407.  
XX PR 04-MAY-1999; 99US-0133484.  
XX PR 05-MAY-1999; 99US-0133485.  
XX PR 06-MAY-1999; 99US-0133486.  
XX PR 06-MAY-1999; 99US-0133487.  
XX PR 07-MAY-1999; 99US-0133863.  
XX PR 11-MAY-1999; 99US-0134256.  
XX PR 14-MAY-1999; 99US-0134218.  
XX PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158332.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159320.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 11.4%; Score 96.5; DB 21; Length 408;  
Best Local Similarity 25.0%; Pred. No. 0.58;  
Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;

QY 10 KDTGVSELRP-HRTVTITQNGKMSSTIVSEEDFLPYKQBLEKGYQPDGMEISGFEG 68  
DB 197 KNKKKEDPLKPGHSAFLVYANERPAALRENRKSVYEVAK-----ITGEEMKNTSD 248  
QY 69 KKDAGY--VINSKOTFIRPVFKTIEKKEENKPTFDVSKKKDNPQVNHQSOLNESHKKE 126  
DB 249 KKKAFYEKAKKKKETYLD-AMEEYKRTKEE-----ALSÖKKE-----FEELIKLHKE 297  
QY 127 DLQREHSGQSDSTKDVATVLDKNNISKSSTTN--PNK 164  
DB 298 ALQMLKKKKEKTDN-----LIKKEKATKKKKNNENVDPNK 330

## RESULT 15

AAG37133  
ID AAG37133 standard; Protein; 456 AA.

AC AAG37133;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 45609.

KM Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137702.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.

	PR	25-OCT-1999;	99US-0161404.
	PR	25-OCT-1999;	99US-0161405.
	FR	25-OCT-1999;	99US-0161406.
	PR	26-OCT-1999;	99US-0161359.
	PR	26-OCT-1999;	99US-0161360.
	PR	26-OCT-1999;	99US-0161361.
	PR	28-OCT-1999;	99US-0161920.
	PR	28-OCT-1999;	99US-0161992.
	PR	28-OCT-1999;	99US-0161993.
	PR	29-OCT-1999;	99US-0162142.
	Query Match	11.4%:	Score 96.5; DB 21; Length 456;
	Best Local Similarity	25.0%:	Pred. No. 0.67; Mismatches 60; Indels 31; Gaps 8
	Matches 40;	Conservative 29;	Mismatches 60; Indels 31; Gaps 8
Oy	10	KDTGEVSELRK-P-HRVYVTIQNGKENSSTIVSEDFILPVYKGELEKGQFDGWEISGFEG	68
Dd	245	KNKKEKDPLKPHPVSAPFLVYANERRALREKNSVVEVAK-----ITGEEMKVLSID	296
Oy	69	KDAGI--VINISKDTFIKVPVKKIIEKKKEENKPTPDVSKKKDPQVNHQSINSHRKE	126
Dd	297	KKKAPPEYAKKDKKEETYLQ-AEEYKYRTKEE-----ALSQKE-----BEELKLKHQE	345
Oy	127	DLGREHSSOKSDSTKVATATVLDKNNISKSSTNN--PNK	164
Dd	346	ALQMFKKEKTQN-----LTKKEKAYKKKKKENVDENPK	378
RESULT 16			
ID	AAG37132	standard; Protein; 476 AA.	
XX	AAG37132;		
AC	18-OCT-2000	(first entry)	
DT			
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 45608.		
DE			
XX	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
XX	EP1033405-A2.		
PN			
XX	06-SEP-2000.		
PD			
XX			
PF	25-FEB-2000;	2000EP-0301439.	
XX			
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0128845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
TR	06-MAY-1999;	99US-0132487.	
TR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	

PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147304.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154033.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155133.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160747.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 11.4%; Score 96.5; DB 21; Length 476;  
 Best Local Similarity 25.0%; Pred. No. 0.71;  
 Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;

QY 10 KDTGEVSEIKP-HRVTVTIQNGKMSSTIVSEDFILPYKGELEKGYQPDGMEISGFEFG 68  
 DB 265 KNKKEKQPLKPKPVSFAFLVYANERBALREKKSVEYVAK-----ITGEMKQLSD 316  
 QY 69 KKDAGI--VINLSKDTPIKPFVKIKIEKKEENKPFVDVSKKDNQVNHSQLNESHKE 126  
 DB 317 KKKAPKEXYAKKKNKTYLQ-AMEEYKRTKEE-----ALSQKE-----EEELKLKQOE 365  
 QY 127 DLQREHSOKSDSTKDVATVLDKNNISSKSTNN--PNK 164  
 DB 366 ALQMLKKKEKTDN-----LIKKEKATKKKKKNENVDPNK 398

RESULT 17  
 AAE20967  
 ID AAE20967 standard; Protein; 2060 AA.

XX AAE20967;  
 AC  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus lugdunensis von Willebrand factor binding protein.  
 KW Von Willebrand factor binding protein; vwb; immunogen; antibacterial;  
 KW vaccine; infection.  
 XX  
 OS Staphylococcus lugdunensis.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..47  
 FT Protein /label= Signal\_peptide  
 FT 48..2060  
 FT /note= "Mature von Willebrand factor binding protein"  
 XX  
 PN WO200228892-A1.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 06-APR-2001; 2001WO-SE00766.  
 XX  
 PR 04-OCT-2000; 2000SF-0003573.  
 XX  
 PA (BIOS-) B10STAPRO AB.  
 XX  
 PI Guss B, Frykberg L, Jacobsson K, Ahlen J, Nilsson M;  
 XX WPI; 2002-304928/34.  
 DR N-PSDB; AAD3371.  
 XX  
 PT New von Willebrand factor binding protein from Staphylococci, useful  
 PT for determining and treating staphylococcal infection -  
 XX  
 PS Claim 3; Page 35-41; 53pp; English.  
 CC The present invention relates to von Willebrand factor binding protein or  
 CC polypeptide (vwb) from Staphylococci. The vwb and immunogens of vwb are  
 CC useful in vaccines to combat infections caused by Staphylococci. The

CC invention is also useful for detection of staphylococcal infection and  
 CC purifying von Willebrand factor from a complex solution. The present  
 CC sequence is Staphylococcus lugdunensis vwb1 protein.

SQ Sequence 2060 AA;

Query Match 11.4%; Score 96.5; DB 23; Length 2060;  
 Best Local Similarity 25.7%; Pred. No. 5.5;  
 Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;

QY 14 EVSEIKPRAVTVTIQ-NGKMSSTIVSEDFILPYKGELEKGYQPDG--WEISGFEK 70  
 DB 1914 DIDEIK--ITIVVDINGRE-----IVPSRKQQLP-EDFIGDMQYTHK-- 1955  
 QY 71 DAGVIINLSKDTPIKPFVKIKIE-----KKEENKPFVDVSKKDNQVNHSQLNESH 123  
 DB 1956 -----IKDDITTYIKKVNNAVPAKQKTKTN--TOSBSQKHPQVQOLVKYN 2006  
 QY 124 RKE--DLQREHSOKSDSTKDVATVLDKNNI 153  
 DB 2007 VKEQRSIEKSHITDHSVSELPEGTANNGL 2038

RESULT 18  
 AAB18272  
 ID AAB18272 standard; Protein; 2500 AA.

XX AAB18272;  
 AC  
 DT 07-NOV-2000 (first entry)  
 XX  
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:129.  
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
 KW antimalarial; malaria; protozoacide; infection; insecticide.  
 XX  
 OS Plasmodium falciparum.  
 XX  
 PN WO200025728-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 05-NOV-1999; 99WO-US26796.  
 XX  
 PR 05-NOV-1998; 98US-0107131.  
 XX  
 PA (HOFF/) HOFFMAN S.  
 PA (CARU/) CARUCCI D.  
 PA (GARD/) GARDNER M.  
 PA (VENT/) VENTER J C.  
 XX  
 PI Hoffman S, Carucci D, Gardner M, Venter JC;  
 XX WPI; 2000-365347/31.  
 DR  
 XX  
 PT Proteins encoded by chromosome 2 of the human malarial parasite,  
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
 PT diagnosis of P.falciparum infection -  
 XX  
 XX Disclosure; Page 302-309; 577pp; English.  
 XX  
 CC The present invention describes proteins and their fragments (I) encoded  
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
 CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
 CC vaccines against P. falciparum infection comprising (I) or (II).  
 CC (I) and (II) are useful for the development of vaccines against  
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
 CC antibody raised to immunogens comprising the sequences of (I), are  
 CC useful in the detection of infection with P. falciparum. Furthermore,  
 CC (I) (especially when they are rifins or secreted or membrane proteins)  
 CC can aid the identification of drugs to treat or prevent P. falciparum  
 CC infection, or they can be used to identify drug resistance in  
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the



CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AA70078 to AA70287 and AB18144 to AB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.

XX Sequence 2500 AA;

Query March 11.3%; Score 96; DB 21; Length 2500;

Best Local Similarity 26.1%; Pred. No. 8.1; Mismatches 54; Indels 32; Gaps 7;

Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;  
19 KPHRVTVTTONGKMSSTVSEDFILPYKGELEKGYQPDGWEISGEGKKAGYVINTL 78  
2173 KPYKIT---ENNNK-----NEGNEILKKYSIENEKNNNYDEQENECILDKDTQCNNVT 2223

QY 79 -----SKDTFIKVFVKKIEKKKEENKPTFDVSKKDNPOVNSQLNE-----SHRKED 127

Db 2224 KEKNNLDNKKSPSPNISKYLEEEKSDDKRD---DKKNDNTEKNNLDNKKSPSPNISKY 2280

QY 128 LOREHSQKSDSTKDVTAIVLDKNNISSKSTNNPNK 164

Db 2281 LEEB---KSDDKRD-----DKKNDNTEKNNLDNKK 2308

RESULT 19

ABJ19106 ABJ19106 standard; Protein; 645 AA.

XX AC ABJ19106;

DT 06-MAR-2003 (first entry)

DE Pathogen specific antigen related staphylococcal protein SEQ ID No 414.

XX Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;  
XX hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
XX auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
XX autoimmune disease; HIV; hepatitis.

XX Staphylococcus sp.

OS WO200259148-A2.

PN 01-AUG-2002.

PD 21-JAN-2002; 2002WO-BP00546.

PR 26-JAN-2001; 2001AT-0000130.

PA (CISTEM-) CISTEM BIOTECHNOLOGIES GMBH.

XX Weinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W;  
XX Minh B, Vytvytska O, Etz H, Dryla A, Weichart T, Hafner M;  
XX Tempelmaier B;

PI WPI; 2003-075410/07.

DR WPI; 2003-075410/07.

XX Identifying, isolating and producing hyperimmune serum-reactive  
XX antigens from a pathogen, for preparing vaccine or medicament for  
XX treating or preventing e.g. staphylococcal infections, comprises  
XX providing antibody preparation

PS Example 7; Page 220; 252pp; English.

XX The invention relates to a novel method for identifying, isolating and  
XX producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
XX allergen, a tissue or host prone to auto-immunity, where the antigens  
XX are used in a vaccine, comprises providing antibody preparation from a

CC plasma pool of a type of animal, or individual sera with antibodies  
CC against the specific pathogen, tumour, allergen, tissue or host prone to  
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of  
CC the 62 sequences of 53-2261 amino acids fully defined in the  
CC specification, or their hyperimmune fragments are useful for the  
CC manufacture of a pharmaceutical preparation, particularly a vaccine  
CC against staphylococcal infections or colonisation against S. aureus or S.  
CC epidermidis. The preparation of antibodies is useful for the manufacture  
CC of a medicament for treating or preventing staphylococcal infections or  
CC colonisation against S. aureus or S. epidermidis. The antibody  
CC preparations may also be used for diagnostic and imaging purposes. Other  
CC conditions that can be treated include cancer, autoimmune diseases or  
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or  
CC protozoan pathogens. This sequence represents a staphylococcal protein  
CC relating to the method for identifying and producing pathogen specific  
CC antigens of the invention.

XX Sequence 645 AA;

Query March 11.0%; Score 93.5; DB 24; Length 645;

Best Local Similarity 21.9%; Pred. No. 2.2; Mismatches 74; Indels 67; Gaps 11;

Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

QY 1 TTVKEFLNKDTGGE-VSELKPHRVTV--TIONGKMSSTVSEDFILPYKGELE-KGY 56

Db 327 SAITEFQNVQPTNKKMTDLQDTKYVYVESVNNNSMDTFYKH-----PIYGMINGKGY 381

QY 57 QF-----DGWEISGFBGK-----D 71

Db 382 MWMETNDNDYKMDPMVEGQVRVITISKAKNNRTIIPYVKGKTLYDAIVKHTIDYD 441

QY 72 AGVYINL-SKDTFIKVFVKKIEKKKEENKPTFDV--SKKDNPOVNSQLNESHK 125

Db 442 GQYHVRIVDKKAFKTAINTDKSNKKEQDQNSAKKATPATPSKPTSPVKEKSDQSDQKD 501

QY 126 EDLQ---REHSQKSDSTKDVTAIVLDKNNISSKSTNNPNK 164

Db 502 DNKQPSVKEKNDASSBSGKDTPATKPTKQEVSSST--PTK 543

RESULT 20

ABP56879 ABP56879 standard; Protein; 654 AA.

XX AC ABP56879;

DT 07-APR-2003 (first entry)

DE Staphylococcus epidermidis KrtN protein SEQ ID NO.10.

XX Staphylococcus aureus; Staphylococcus epidermidis; MSCRAMM; antibody;  
XX microbial surface component recognising adhesive matrix molecule;  
XX surface protein; infection; antibacterial; antiinflammatory; vaccine;  
XX immunosuppressive; antirheumatic; gene therapy; pneumonia; endocarditis;  
XX septic arthritis; biomaterial related infection.

XX Staphylococcus epidermidis.

OS WO2002102829-A2.

PN 27-DEC-2002.

PD 17-JUN-2002; 2002WO-US19220.

PR 15-JUN-2001; 2001US-298098P.

PA (INH-) INHIBITEX INC.

PA (QUB-) QUBEN ELIZABETH COLLEGE DUBLIN.

PI Foster TJ, Roche F, Patti JM, Hutchins JT, Hall A, Domanski P;  
PI Patel P, Stribeys P, Speziale P;

XX



AC AAG82397;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
XX S. epidermidis open reading frame protein sequence SEQ ID NO:1898.  
DE Staphylococcus epidermidis SRI strain; infection; diagnosis;  
XX vaccination; endocarditis.  
XX  
OS Staphylococcus epidermidis.  
XX  
FN MO200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000MO-US30782.  
XX  
PR 09-NOV-1999; 99US-0164258.  
XX  
PA (GLAX ) GLAXO GROUP LTD.  
XX  
P1 Kimmery MJ;  
XX  
DR WPI; 2001-316495/33.  
DR N-PSDB; AAH53247.  
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX  
PS Claim 18; Page 516; 2188bp; English.  
XX  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce host cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX  
SQ Sequence 299 AA;  
Query Match 10.8%; Score 92; DB 22; Length 299;  
Best Local Similarity 24.4%; Pred. No. 1;  
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;  
OY 9 NKDTGEVSELKPHRTVTYIIONGKEMSTIVSEEDFI---LPVYGELEKG-YOPDGWEI 63  
DB 33 NKDT-EKSDKTKYHRIISLPSTELVRLGIGEDIVGSTVDYDKDKKKQGDANML 91  
OY 64 SGFE-----GKDDAGVYINLSKDTFIKPV-----FKKIEKK 95  
DB 92 NKEELIKAKPDLILAHESQKNSAGKVLKSLDKGVVVVYVDAQSIDETDTPFKSIGOLT 151  
OY 96 EENKPTFDVSKKKDNPQVNSQLNESHKEDLQREHSQ 135  
DB 152 DREKQAKELVDETKEINVEKIINSVPKHHKKOEVEFMEVSSK 191  
RESULT 23  
ABP40822

ID ABP40822 standard; Protein; 309 AA.  
XX  
AC ABP40822;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5667.  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5667.  
XX  
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
FN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
DR N-PSDB; ABN93367.  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 5667; 267bp; English.  
XX  
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP3124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 309 AA;  
Query Match 10.8%; Score 92; DB 23; Length 309;  
Best Local Similarity 24.4%; Pred. No. 1.1; 60; Indels 34; Gaps 5;  
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;  
OY 9 NKDTGEVSELKPHRTVTYIIONGKEMSTIVSEEDFI---LPVYGELEKG-YOPDGWEI 63  
DB 43 NKDT-EKSDKTKYHRIISLPSTELVRLGIGEDIVGSTVDYDKDKKKQGDANML 101  
OY 64 SGFE-----GKDDAGVYINLSKDTFIKPV-----FKKIEKK 95  
DB 102 NKEELIKAKPDLILAHESQKNSAGKVLKSLDKGVVVVYVDAQSIDETDTPFKSIGOLT 161  
OY 96 EENKPTFDVSKKKDNPQVNSQLNESHKEDLQREHSQ 135  
DB 162 DREKQAKELVDETKEINVEKIINSVPKHHKKOEVEFMEVSSK 201  
RESULT 24  
ABP38188  
ID ABP38188 standard; Protein; 442 AA.  
XX  
AC ABP38188;  
XX  
DT 24-JUL-2002 (first entry)  
XX

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3033.  
XX  
KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
DR N-PSDB; ABN90733.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 3033; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 442 AA;  
XX  
Query Match 10.7%; Score 91; DB 23; Length 442;  
Best Local Similarity 21.7%; Pred. No. 2.2;  
Matches 49; Conservative 30; Mismatches 65; Indels 82; Gaps 11;  
QY 10 KDTGEVSELPKRVVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK-----54  
DB 90 EDT-EINPVAQPEVNTQLEKQDFEATVVEPEVKLGDKLEIEKQETDLSDEIQ 148  
QY 55 -----GVQFDGMEISG--FEQKXAGYVINSKDTFTK 85  
DB 149 ESIDHSLSLAEVNVKEDGAVNGDTVINIDFSG-SVDGEFFDGGQAGEYDLEIGSGSFLP 207  
QY 86 PVFKKIEEKKKEENK-----PTFDVS---KKKNPQVNHNSOLNE- 121  
DB 208 GFEEQIEGKTKGEKQVTVTFPEEYHAEELAGKEATFKTKVNIKFKQVPELNDIELNEL 267  
QY 122 -----SHRKEDLQREHSHQSKSDSTKDVATVATVLDKNINSKSTNN 160  
DB 268 DSDAENVDEYKENLRKRLSEOKATEAMNT-----EKKEALNKATEN 308  
XX  
RESULT 25  
ABP40312  
ID ABP40312 standard; Protein; 902 AA.  
XX  
AC ABP40312;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5157.  
XX

KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
DR N-PSDB; ABN92857.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 5157; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 902 AA;  
XX  
Query Match 10.7%; Score 91; DB 23; Length 902;  
Best Local Similarity 24.9%; Pred. No. 6.1;  
Matches 43; Conservative 31; Mismatches 75; Indels 24; Gaps 8;  
QY 1 TTYKPEILNKDGEVSELPKRVVTVTQNGKMS-STIVSEEDFILPVYKGLEKGYQPD 59  
DB 736 TPVSEYRLSNRQGGKIK---TATYTERGNITVCTTVGEBDLNVVNAVY---IRLD 788  
QY 60 GWEISGFESKXQAGY-VINISKDTFTKPVFKKIEEKKKEEN-----KPTFDVSKKK 109  
DB 789 VHDISQ-NGRAAQGYRLMKLGCGQFVSTVAKXNEEDNENMDAQAQSTTETADVEEV 847  
QY 110 DNPQVNHNSQLNSHRKEDLQREE-HSQKSDSTKDVATVATVLDKNINSKSTNN 161  
DB 848 D-----DQTPGNAIHTEGDAEMESVESPENDDRIDIROPFMDRVNEDIBESASDN 896  
XX  
RESULT 26  
AAB42897  
ID AAB42897 standard; Protein; 209 AA.  
XX  
AC AAB42897;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF2661 polypeptide sequence SEQ ID NO:5322.  
XX  
XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;  
XX vulnerable; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;  
XX anticonvulsant; osteoplastic; antiarthritic; immunosuppressant; cardiant;  
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
XX hypotensive; dermatological; immunosuppressive; antiinflammation;  
XX antiviral; antibacterial; antifungal; antineumatic; antihypoid;



Matches	38; Conservative	40; Mismatches	70; Indels	23; Gaps	8; ...
Oy	6	FLNKDGTGEVSELKPH---RVTVTIQLNGKEM-----SSTIVSEDDFLIPYKGLERKY	56		
Db	64	YTANKE--KMKKFKSPWTRMPFTRPKQGIELNHVWVKGSKELIEQGF-----EEDGTPKPY	117		
Oy	57	QPDGW---ETSGF--EGKKAGGYINLSK-DTFLKPVFKKIEKKKEENKPTPDVSKK	109		
Db	118	FFEKKNVQLLEPFVDEDDTYDL-VWIEIKVEYSKKKEERAREKREKREKRDLEKKK	176		
Oy	110	DNPQVNSQLNESHRRKDLQREHESQKSDSTKDTATVLDKNNISKSTTN	160		
Db	177	QQQQQKSKQQNPQNGIKDDEKNQDTRNNTKDKDSQKSEKDPVTYAKKETD	227		
RESULT 28					
ID	AAV44364 standard, protein; 757 AA.				
AC	AAV44364;				
DT	14-MAR-2000 (first entry)				
DE	Human cell cycle regulation protein-5.				
XX					
XX	CECRP-5; cell cycle regulation protein-5; cell proliferation;				
KM	cell proliferative disease; cancer; atherosclerosis; cirrhosis;				
KM	hepatitis; psoriasis; immune system disorder; allergy; asthma;				
KM	acquired immune deficiency syndrome; Crohn's disease; Blast method;				
KX	rheumatoid arthritis; gene therapy; chromosomal mapping.				
XX					
OS	Homo sapiens.				
XX					
Key	Location/Qualifiers				
PH	4				
FT	Modified-site				
FT	/note= "Potential glycosylation site"				
FT	63				
FT	Modified-site				
FT	/note= "Potential glycosylation site"				
FT	479				
FT	Modified-site				
FT	/note= "Potential glycosylation site"				
FT	498				
FT	Modified-site				
FT	/note= "Potential glycosylation site"				
FT	31				
FT	Modified-site				
FT	/note= "Potential phosphorylation site"				
FT	43				
FT	Modified-site				
FT	/note= "Potential phosphorylation site"				
FT	81				
FT	Modified-site				
FT	/note= "Potential phosphorylation site"				
FT	96				
FT	Modified-site				
FT	/note= "Potential phosphorylation site"				
FT	99				
FT	Modified-site				
FT	/note= "Potential phosphorylation site"				
FT	130				
FT	Modified-site				
FT	/note= "Potential phosphorylation site"				
FT	158				
FT	Modified-site				
FT	/note= "Potential phosphorylation site"				
FT	162				
FT	Modified-site				
FT	/note= "Potential phosphorylation site"				
FT	163				
FT	Modified-site				
FT	/note= "Potential phosphorylation site"				
FT	184				
FT	Modified-site				

FT	Modified-site	/note= "Potential phosphorylation site"
FT	323	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	343	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	367	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	370	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	432	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	488	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	499	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	207	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	277	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	298	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	370	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	419	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	474	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	488	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	499	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	528	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	532	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	539	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	609	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	624	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	730	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
XX		
PN	W09964593-A2.	
PD	16-DEC-1999.	
XX		
PF	08-JUN-1999;	99WO-US12906.
XX		
PR	08-JUN-1998;	98US-0088695.
XX		
PA	(INCY-) INCYTE PHARM INC.	
XX		
PI	Bandman O, Lal P, Tang YT, Corley NC, Guejler KJ, Baughn MR, Pactereson C,	
XX		
DR	WPI; 2000-105887/09.	
XX		
DR	N-PSDB; AAZ29484.	
XX		
PT	Novel regulatory proteins, for diagnosis, treatment and prevention of cell proliferative and immune system diseases	
XX		
PS	Claim 1; Page 70-71; 87pp; English.	
XX		
CC	The present sequence is cell cycle regulation protein-5 (CECRP-5). Blast	
CC	analytical method was used to identify this protein. CECRPs are	
CC	activators of cell proliferation or inhibitors of cellular processes that	
CC	modulate proliferation. They are used to treat or prevent cell	
CC	proliferative diseases like cancers, atherosclerosis, cirrhosis,	
CC	hepatitis, psoriasis, immune system disorders (e.g. acquired immune	
CC	deficiency syndrome, allergy, asthma, Crohn's disease, rheumatoid	
CC	arthritis). Antibodies are raised to screen for specific binding agents.	

CC The corresponding nucleic acid is used in gene therapy, chromosomal  
CC mapping and isolation of related sequences.

XX Sequence 757 AA;

Query Match 10.7%; Score 90.5; DB 21; Length 757;

Best Local Similarity 28.1%; Pred. No. 5.3;

Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 13 GEVSELKPHRYVTITQNGKMSSTIVSE--EDFLIPYKGELEKGYQF----DGWE--- 62

DB 398 GELQOLSSGSQI-----HGKSDSPNVYTEKEIALLRELTLEKLTFTQQRSDIMERLY 452

QY 63 -----ISGFEKKDAG---YVINLSKDTFIPVKKLEKKEENKPTFVSKKXD 110

DB 453 VEAKDQNGKQGTGDKKKGGSHRAKKSKETFLGSV-----KETFDAMKST 500

QY 111 NPQVNH-----SQLNESHKEDLQREHSQKSD--STKDVTATVLDK 150

DB 501 KEFVHHHEKIKQAKEA-VKENLKKFSDSVKSTFRHFDTTKNIFDE 546

#### RESULT 29

AAB53319

ID AAB53319 standard; Protein; 758 AA.

AC AAB53319;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen protein sequence SEQ ID NO:859.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW identification; cytostatic; cardioactive; neuroprotective; vulnery;

KW immunomodulatory; muscular; gynaecological; gastrointestinal;

KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;

KW neutral disorder; immune system disorder; muscular disorder;

KW reproductive disorder; gastrointestinal disorder; renal disorder;

KW infectious disease; cardiovascular disorder.

OS Homo sapiens.

XX WO200055351-A1.

XX 21-SEP-2000.

PD 08-MAR-2000; 2000MO-US05883.

XX 12-MAR-1999; 99US-0124270.

PR (HUMA-) HUMAN GENOME SCI INC.

XX PA ROSEN CA, RUBEN SM;

XX PI WPI; 2000-587534/55.

XX DR N-PSDB; AAC98076.

XX Colon cancer associated gene sequences, referred to as colon cancer

PT antigens, useful for the treatment, prevention, and diagnosis of colon

XX disorders such as colon cancer -

PS Claim 11; Page 1408-1411; 2104pp; English.

CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,

CC called human colon cancer antigens, given in AAB53319 to AAB54006. The

CC human colon cancer antigens can have cytostatic, cardioactive, muscular;

CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,

CC vulnery, nephrotropic, antiinfective and antibacterial activities, and

CC can be used in gene therapy. The colon cancer antigen polynucleotides,

CC proteins and antibodies to the proteins are useful for the prevention,

CC treatment and diagnosis of colon disorders, such as colon cancer. The

CC polynucleotides may be used in diagnostics and research, such as for

CC chromosome identification, and as hybridisation probes. The proteins

CC may also be used to prevent diseases such as neural disorders, immune  
CC system disorders, muscular disorders, reproductive disorders,  
CC gastrointestinal disorders, wounds, renal disorders, infectious  
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
CC AAB54007 represent sequences used in the exemplification of the present  
CC invention.

XX Sequence 758 AA;

Query Match 10.7%; Score 90.5; DB 21; Length 758;

Best Local Similarity 28.1%; Pred. No. 5.3;

Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 13 GEVSELKPHRYVTITQNGKMSSTIVSE--EDFLIPYKGELEKGYQF----DGWE--- 62

DB 399 GELQOLSSGSQI-----HGKSDSPNVYTEKEIALLRELTLEKLTFTQQRSDIMERLY 453

QY 63 -----ISGFEKKDAG---YVINLSKDTFIPVKKLEKKEENKPTFVSKKXD 110

DB 454 VEAKDQNGKQGTGDKKKGGSHRAKKSKETFLGSV-----KETFDAMKST 501

QY 111 NPQVNH-----SQLNESHKEDLQREHSQKSD--STKDVTATVLDK 150

DB 502 KEFVHHHEKIKQAKEA-VKENLKKFSDSVKSTFRHFDTTKNIFDE 547

#### RESULT 30

ABP68963

ID ABP68963 standard; Protein; 801 AA.

AC ABP68963;

DT 20-JAN-2003 (first entry)

DE Human polypeptide SEQ ID NO 1010.

XX Human; genome mapping; gene therapy; food supplement; virus; fungus;

KW cell-proliferative disorder; neurodegenerative disease; bacterial;

KW Parkinson's disease; Alzheimer's disease; autoimmune disease;

KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;

KW arthritis; cytostatic; immunomodulator; nocotropic; neuroprotective;

KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;

KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;

KW antiarthritic.

OS Homo sapiens.

XX WO200270539-A2.

XX 12-SEP-2002.

PD 05-MAR-2002; 2002WO-US05095.

XX 05-MAR-2001; 2001US-0799451.

XX (HXSE-) HXSEQ INC.

XX PA Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;

XX PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;

XX PI Wehman T, Wang J, Wang D, Drmanac RT;

XX WPI; 2002-759812/82.

XX DR N-PSDB; AB211180.

XX New polynucleotides comprising sequences assembled from expressed

PT sequence tags (ESTs), useful for treating cell-proliferative,

PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or

PT platelet or coagulation disorders -

PS Claim 9; SEQ ID NO 1010; 1012pp + Sequence Listing; English.

CC The invention relates to an isolated polynucleotide (I) comprising a

CC nucleotide sequence selected from any of 948 sequences

(AB21119-AB212066) or their mature protein coding portion, active domain CC coding protein or complementary sequences. The polynucleotides are useful CC for identifying expressed genes or for physical mapping of human genome. CC The encoded polypeptides (AB68902-AB69849) are useful as molecular CC weight markers, as a food supplement, for generating antibodies, in CC medical imaging, screening and diagnostic assays and for treating CC cell-proliferative disorders (cancer), neurodegenerative diseases CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid CC disorders, platelet or coagulation disorders, wound, burns, incision, CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal, CC parasitic), arthritis, etc. CC Note: The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 801 AA;

Query Match 10.7%; Score 90.5; DB 23; Length 801;  
Best Local Similarity 22.1%; Pred. No. 5.8; Mismatches 34; Conservative 35; Indels 31; Gaps 6;

DB 4 KEFLNKDTGSEVSELPKRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQPDGWEI 63  
504 KEFLNKDTGSEVSELPKRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQPDGWEI 551

QY 64 SCFEGKKDAGYVNLKDTFIKVFVKIEKKEENKPTFDV---SKKKNQVNHSQL 119  
552 -----DLGQELRLTRFQ-VQNSHTELAERHQVQAOERIEIRLSSELEDMKLSKEXL 603

DB 120 NESHKEDLQREHESQKSDSTKQVTAIVLTKNNI 153  
604 -----EKAKLELE---AQDTVSNLHQOVQVDRNEV 630

QY 604 -----EKAKLELE---AQDTVSNLHQOVQVDRNEV 630

DB 604 -----EKAKLELE---AQDTVSNLHQOVQVDRNEV 630

RESULT 31  
AAG85023  
ID AAG85023 standard; Protein; 922 AA.

AC AAG85023;  
XX  
DT 11-SEP-2001 (first entry)  
XX  
DE Shrimp white spot Bacilliform virus (WSBV) protein 114.  
XX  
KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;  
KM antiviral agent; gene expression; antisense construct;  
XX transgenic viral resistant shrimp.  
XX  
OS White spot syndrome virus.  
XX  
PN WO200136351-A2.  
XX  
PD 31-MAY-2001.  
XX  
PF 08-NOV-2000; 2000WO-US28888.  
XX  
PR 24-NOV-1999; 99CN-0124717.  
XX  
PA (PENY-) PE CORP NY.  
PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.  
PA (SINO-) SINOGENOMAX CO LTD.  
XX  
PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;  
XX  
XX WPI; 2001-355877/37.  
DR N-PSDB; AAH62803.  
XX  
PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus  
PT (WSBV), useful for producing viral polypeptides that can be used to  
PT screen for agents that are useful for treating WSBV infection -  
XX  
PS Claim 1; Figure 3; 626bp; English.

XX The invention provides the primary nucleotide sequence of the WSBV genome CC (AA62689), predicted transcript sequences (AA62689-AAH62839) and CC encoded proteins (AA684910-AA685051) and oligonucleotide sequences CC (AA62840-63160) suitable for use as primers or probes. The nucleic acid CC molecules and proteins of the invention are useful for diagnosis and CC monitoring viral infection, in screens for antiviral agents and for CC monitoring viral gene expression or activity during a treatment regimen. CC The nucleic acid molecules are also useful as antisense constructs to CC control viral gene expression in infected cells and tissues and to create CC transgenic viral resistant shrimp.

XX Sequence 922 AA;

Query Match 10.7%; Score 90.5; DB 22; Length 922;  
Best Local Similarity 22.0%; Pred. No. 7; Mismatches 42; Conservative 30; Indels 68; Gaps 8;

QY 10 KDTGEVSELPKRVTVTIQNG-----KMSSTIVSEDFILPVY---- 48  
300 KTTNQIVE---BSVYESRNGCCCFKNDKWLAKRSNLKSLNNTVFGEDDEKSAVAYS 356

DB 49 -----KGELEKGYQPDGWEISGFEKKDAGYVNLKDTFIKVF-FKIEKKEENK 100  
357 SEDEDEDENEBEVDYDYNNETIESSVG-----NVIKNLIRKTIQGLDVEEKEEGEQ 408

QY 101 PFVDYSKKKNQVNHSQLNESHKEDL-----QREHSQKSDSTKQVTAIVL----DK 150  
409 SEEEEDSDDDDDASSVSSSSSSSVTVVAALAEEDDEDEKOKD-TATVEDEDDK 467

DB 151 NNISKSTTN 161  
468 ESVISSSSSDS 478

QY 151 NNISKSTTN 161  
DB 468 ESVISSSSSDS 478

RESULT 32  
ABB58769  
ID ABB58769 standard; Protein; 1183 AA.

AC ABB58769;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 3099.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL02872.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 3099; 21bp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent



CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB101840-AB130511), expressed DNA  
 CC sequences (AB57737-AB872072).  
 CC (AB57737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 1183 AA;

Query Match 10.6%; Score 90; DB 22; Length 1183;

Best Local Similarity 24.7%; Pred. No. 11; Mismatches 54; Indels 33; Gaps 5;

Matches 36; Conservative 24; Mismatches 54; Indels 33; Gaps 5;

CC 19 KPHRTVTITQNGKMSSTIVSEDFLLPYKGELEKGYQPDGMEISGFEKGADAGVINV 78

CC 310 KPEVTTAKDKNKD---PLIADSKPKVPKKTETSK-----TGKKGKKEGA----- 352

CC 79 SKDTFKPVFKKIEKKKEENKPTDVSKKKONPOVNHSQLNESHKEDLQREHSQ--- 135

CC 353 -----KPAEK--SKKEEKEDSTSKSKKKADSPANNQKEQIGVKKTSBEDASNSHK 404

CC 136 ----KSDTKDVTATVLDKNISSKS 157

CC 405 SVTAKKDSKKDEASTQVKSNDNSPEA 430

RESULT 33

AAV03190

ID AAV03190 standard; Protein; 402 AA.

XX AAV03190;

DT 16-JUN-1999 (first entry)

XX S. aureus trigger factor protein sequence.

XX Trigger factor; tig; diagnosis; immune response; bacterial infection;

XX S. aureus infection; otitis media; toxic shock syndrome; conjunctivitis;

XX wound infection; impetigo; septic arthritis; therapy.

XX Staphylococcus aureus.

OS Staphylococcus aureus.

XX EP905233-A2.

PN EP905233-A2.

XX 31-MAR-1999.

PD 31-MAR-1999.

XX 21-AUG-1998; 98EP-0306697.

PF 21-AUG-1998; 98EP-0306697.

XX 29-DEC-1997; 97US-0999339.

PR 04-SEP-1997; 97US-0057511.

XX (SMK ) SMITHKLINE BEECHAM CORP.

PA (SMK ) SMITHKLINE BEECHAM CORP.

XX Burnham MKR, Fosberry A, Hodgson JE, Jaworski JD;

PI Lawlor EJ, Rosenberg M, Traini CM, Wang M, Ward J;

PI Warren RL;

XX WPI; 1999-192663/17.

DR N-PSDB; AAX28175.

XX New Staphylococcus aureus trigger factor (tig) polypeptide and

PT polynucleotide, useful as diagnostic reagents and for prevention and

PT treatment of Staphylococcus aureus infections

XX Claim 23; Page 8; 31pp; English.

PS This sequence is the Staphylococcus aureus trigger factor (tig)

CC polypeptide of the invention. Tig polypeptides and polynucleotides are

CC useful for diagnosing diseases related to over or underexpression of tig  
 CC protein by identifying mutations in the tig gene, or determining tig  
 CC polypeptide or mRNA expression levels due to an infection of an organism  
 CC with the tig gene. They can diagnose the stage and type of infection. Tig  
 CC polypeptides are also useful for screening compounds which affect  
 CC activity of the protein by measuring the binding to tig and observing the  
 CC stimulation or inhibition of the polypeptide function. These can be used  
 CC in treatment to inhibit or enhance tig activity, in addition to direct  
 CC administration of tig polypeptides to treat conditions associated with a  
 CC lack of tig polypeptide, or direct administration of antisense sequences  
 CC to prevent expression. Tig polypeptides (administered directly, in a  
 CC vector and as a vaccine) and antibodies induce an immune response to  
 CC immunise and prevent disease. Diseases diagnosed, prevented or treated  
 CC include: bacterial infections, especially *Staphylococcus aureus*  
 CC infections which cause otitis media, toxic shock syndrome,  
 CC conjunctivitis, wound infection, impetigo and septic arthritis etc. Tig  
 CC polypeptides, polynucleotides and their (ant)agonists can prevent  
 CC adhesion of bacteria to matrix proteins, and are useful for use on wounds  
 CC and body implants to prevent bacterial infection.

CC Sequence 402 AA;

Query Match 10.6%; Score 89.5; DB 20; Length 402;

Best Local Similarity 22.1%; Pred. No. 2.8; Mismatches 60; Indels 83; Gaps 13;

Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

CC 11 DTGEVSELKPHRYVT-IONKE--MSSTIVSEDFLLPYKGELEK----- 54

CC 53 DETDIKPYAOPESVTOIEKDPFEATVPEPVKIGDYKGLIEKQETELSDDELOE 112

CC 55 -----GYQPDGMEISG--FEKGADAGVINVLSKDTFKP 86

CC 113 AIDHSLGHAEVWVEDGVENGDTVINDFSG-SVDGEFEGGQAEGLDLEGGSSFI-P 170

CC 87 VFK-----KIEKKE-----EENKPTPDVS---KKONPOVNHSQLNE- 121

CC 171 GFEQLEGMKVDEKDVVTFPEEYHAEELGKAEFTFKVNEIFKGVPELTDIEIANEL 230

CC 122 -----SHRKDLQREHSQKSDTKDVTATVLDKNISSKSTIN 160

CC 231 DAEANTVDEYKENLKRILAEQKATDAENV-----EKEBAITKATDN 271

RESULT 34

AAV35091

ID AAV35091 standard; Protein; 511 AA.

XX AAV35091;

DT 13-SEP-1999 (first entry)

XX Chlamydia pneumoniae transmembrane protein sequence.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

XX Chlamydia pneumoniae.

OS Chlamydia pneumoniae.

XX WO9927105-A2.

PN WO9927105-A2.

XX 03-JUN-1999.

PD 03-JUN-1999.

XX 20-NOV-1998; 98WO-IB01890.

PF 20-NOV-1998; 98WO-IB01890.

XX 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GSEST ) GENSET.

PA (GSEST ) GENSET.

XX Griffiths R;

PI WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae  
PT  
XX  
PS Page 975-976; Disclosure; 1912pp; English.  
XX  
CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX  
SQ Sequence 511 AA;  
XX  
Query Match 10.6%; Score 89.5; DB 20; Length 511;  
Best Local Similarity 24.5%; Pred. No. 3.9;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
QY 84 IXPVFKIEBKKEENKPTFD-----VSKKQNPQVNHSQLNSHRK 125  
DB 95 VKGVFKKTQAPREVSPLPSHVHGQRLPGLGFRDRIQKSKSENPEDLGMKRSYSD 154  
QY 126 EDIQREHSQKSDTQVATVYDKNNISKSTT 159  
DB 155 GDLDRVGHDSNEDSTEDSRS---EGGEPSKSSS 185  
XX  
RESULT 35  
ABP73570  
ID ABP73570 standard; Protein; 516 AA.  
XX  
AC ABP73570;  
XX  
DT 30-JUN-2003 (first entry)  
XX  
DE Candida albicans essential protein SEQ ID NO 7407.  
XX  
KM Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;  
KM signal transduction; DNA replication; cell division; growth;  
KM proliferation; Candida albicans; fungicide; antifungal.  
XX  
OS Candida albicans.  
XX  
PN WO200253728-A2.  
XX  
PD 11-JUL-2002.  
XX  
PF 26-DEC-2001; 2001WO-US49486.  
XX  
PR 29-DEC-2000; 2000US-259128P.  
PR 20-FEB-2001; 2001US-0792024.  
PR 22-AUG-2001; 2001US-314050P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
XX  
DR WFI, 2002-566694/60.  
XX  
N-PSDB; AB232120.  
XX  
PT Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele  
PT of a gene and placing other allele of the gene under conditional  
PT expression -  
XX  
PS Claim 44; SEQ ID NO 7407; 167bp + Sequence Listing; English.  
XX  
CC The invention relates to constructing (M1) a strain of diploid fungal

CC cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal  
CC cells having both alleles modified are useful for identifying a gene that  
CC is essential to the survival or growth of a fungus, a gene that  
CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
CC that contributes to the resistance of a diploid fungus to an antifungal  
CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthetic, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.  
XX  
SQ Sequence 516 AA;  
XX  
Query Match 10.6%; Score 89.5; DB 23; Length 516;  
Best Local Similarity 23.3%; Pred. No. 3.9;  
Matches 37; Conservative 27; Mismatches 58; Indels 37; Gaps 6;  
QY 10 KDTGEVSELPKPRVTVT--IQNGKEMSPVSEEDFILPVYKGELEKGYQF----- 58  
DB 369 KTKGRISRYLAKKCSIASRIDVYSEPTTAFGE-----ILKKQVEDRLKFTDTSAPMK 422  
QY 59 -----DGEWISGFEKKDAGYVNLKSTFTPKVFKIEBKE---EENKPTF 103  
DB 423 NSDAIKAALALNGQDLGAEQKDVWV---SDEQVKKEKKEKKEKKKKKKKKKK 479  
QY 104 DYSKKKQNPQVNHSQLNSHREKDLQREHSQKSDSTQ 142  
DB 480 DKREKKDKKKKKDKK--DKRRKSDQEEETPKKKKKKKK 516  
XX  
RESULT 36  
AAY03189  
ID AAY03189 standard; Protein; 525 AA.  
XX  
AC AAY03189;  
XX  
DT 16-JUN-1999 (first entry)  
XX  
DE S. aureus trigger factor protein sequence.  
XX  
KM Trigger factor; rig; diagnosis; immune response; bacterial infection;  
KM S. aureus infection; otitis media; toxic shock syndrome; conjunctivitis;  
KM wound infection; impetigo; septic arthritis; therapy.  
XX  
OS Staphylococcus aureus.  
XX  
PN EP905233-A2.  
XX  
PD 31-MAR-1999.  
XX  
PF 21-AUG-1998; 98EP-0306697.  
XX  
PR 29-DEC-1997; 97US-0099339.  
PR 04-SEP-1997; 97US-0057511.  
XX  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PA (SMK ) SMITHKLINE BEECHAM PLC.  
XX  
PI Burnham MKR, Fosberry A, Hodgson JF, Jaworski DD;

PI Lawlor EJ, Rosenberg M, Traini CM, Wang M, Ward J;  
PI Warren RL;  
XX WPI, 1999-192663/17.  
DR N-PSDB; AAX28174.  
XX  
PT New Staphylococcus aureus trigger factor (tig) polypeptide and  
PT polynucleotide, useful as diagnostic reagents and for prevention and  
PT treatment of Staphylococci aureus infections  
XX  
PS Claim 14; Page 6; 31pp; English.  
XX  
CC This sequence is the Staphylococcus aureus trigger factor (tig)  
CC polypeptide of the invention. Tig polypeptides and polynucleotides are  
CC useful for diagnosing diseases related to over or underexpression of tig  
CC protein by identifying mutations in the tig gene, or determining tig  
CC polypeptide or mRNA expression levels due to an infection of an organism  
CC with the tig gene. They can diagnose the stage and type of infection. Tig  
CC polypeptides are also useful for screening compounds which affect  
CC activity of the protein by measuring the binding to tig and observing the  
CC stimulation or inhibition of the polypeptide function. These can be used  
CC in treatment to inhibit or enhance tig activity, in addition to direct  
CC administration of tig polypeptides to treat conditions associated with a  
CC lack of tig polypeptide, or direct administration of antisense sequences  
CC to prevent expression. Tig polypeptides (administered directly, in a  
CC vector and as a vaccine) and antibodies induce an immune response to  
CC immunize and prevent disease. Diseases diagnosed, prevented or treated  
CC include: bacterial infections, especially Staphylococcus aureus  
CC infections which cause otitis media, toxic shock syndrome,  
CC conjunctivitis, wound infection, impetigo and septic arthritis etc. Tig  
CC polypeptides, polynucleotides and their (ant)agonists can prevent  
CC adhesion of bacteria to matrix proteins, and are useful for use on wounds  
CC and body implants to prevent bacterial infection.  
SQ  
SQ Sequence 525 AA;  
Query Match 10.6%; Score 89.5; DB 20; Length 525;  
Best Local Similarity 22.1%; Pred. No. 4;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;  
QY 11 DGEVSELKPHRYVT--IONGKE--MSSITVSEEDFLPYKYG-ELEK----- 54  
DB 81 DETDTPVAVQPEVSVQIQEKQDFEATVVEPEVKLGQYKLEIKQETELSDDELOE 140  
QY 55 -----GYQFDGWEISG--FEGKQAGVYINLSKDTFIK 86  
DB 141 AIDHSLGLHAEVWVEDGVVENDPTVINIDFSG-SVQGEFEGGQAGVYDLIGSGSFI-P 198  
QY 87 VRK-----KIEEKE-----EE--NKPTFDVS---KKKDNPOVNHSQLNF- 121  
DB 199 GFEQLEGKVKVDEKQVVVTFPEEYHAEELAGLEAFKTKVNEIKFEKVEPELDELAINEL 258  
QY 122 -----SHRKEDLQREHSHOKSDSTKDYATVYLDNNISSKSTNN 160  
DB 259 DAEANTVDEYKENLRKRLAEQKATDAENV-----EKEAITYATDN 299  
RESULT 37  
ABBS8704  
ID ABBS8704 standard; Protein; 1408 AA.  
XX  
XX ABB58704;  
AC  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 2904.  
DE  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200171042-A2.

XX  
PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX  
XX 11-JUL-2000; 2000US-0614150.  
PA  
XX  
XX (PEKE ) PE CORP NY.  
PI  
XX  
XX Venter JC, Adams M, Li PMD, Myers EW;  
XX  
XX WPI: 2001-656860/75.  
DR  
XX  
XX N-PSDB; ABL02807.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
PS  
XX  
XX Disclosure; SEQ ID NO 2904; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABBS7737-ABBS72072).  
CC  
XX The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
XX  
SQ  
SQ Sequence 1408 AA;  
Query Match 10.6%; Score 89.5; DB 22; Length 1408;  
Best Local Similarity 24.7%; Pred. No. 16;  
Matches 48; Conservative 23; Mismatches 62; Indels 61; Gaps 9;  
QY 29 NGKE-----MSSITVSEEDF--ILPYKGELEKGYQFDGWEISGFEK--KDAGYV----- 75  
DB 646 NCKQKLNLSKETKTETSKDFKVIKESKE--EKNDQESLELNQKLPKQDDYLDYKRP 704  
QY 76 -----INLSKDTFIKPVFK---KIEEKEENKPTFDVSKKKDNPOVNHSQLN 120  
DB 705 MESKEYSEPLAVKLGEDVSKSPNSSDTKISEVNGEE--PQNGELPRTGQVYQPEKS 762  
QY 121 ESHRK-EDLQREHSHOKSDSTKDYV-----ATVLDX 150  
DB 763 KTQAKTEFLAKVQSELKSKSVQDVTPKEQTKEETHKGSPPREATPKSESPKGTTEVDK 822  
QY 151 NNISKSTNNPNK 164  
DB 823 SNDVSVIKPSEPD 836  
RESULT 38  
ABB93341  
ID ABB93341 standard; Protein; 296 AA.  
XX  
XX ABB93341;  
AC  
XX  
XX 31-MAY-2002 (first entry)  
DT  
XX  
XX Herbicidally active polypeptide SEQ ID NO 2552.  
DE  
XX  
XX Herbicidal; plant; agriculture; herbicide.  
KW Arabidopsis thaliana.  
XX  
XX WO200210210-A2.  
OS  
XX  
XX 07-FEB-2002.

```

XX 28-AUG-2001; 2001WO-EP09892.
XX PF
XX PA
PR 28-AUG-2001; 2001WO-EP09892.
XX (FARB ) BAYER AG.
XX PI
XX PI Tietjen K, Weidler M;
XX MPI; 2002-269010/31.
XX
XX PT Identifying plant target proteins for herbicidally active compounds,
XX PT comprising aligning and comparing nucleic acid or amino acid sequences
XX PT from plant with nucleic acid or amino acid sequences from non-plant
XX PT organisms -
XX PS Claim 5; SEQ ID NO 2552; 261pp + Sequence Listing; English.
XX
XX CC The invention relates to identifying target proteins
XX CC (AB90790-ABB94016) for herbicidally active compounds, comprising
XX CC aligning and comparing nucleic acid or amino acid sequences from plant
XX CC with nucleic acid or amino acid sequences from non-plant organisms using
XX CC suitable search parameters, where plant sequences having an E-value
XX CC greater by a factor of 3 than the E-value of most similar non-plant
XX CC sequences are selected. The polypeptides or nucleic acids encoding them
XX CC are useful for identifying modulators. The identified modulators are
XX CC useful as herbicides.
XX
XX SQ Sequence 296 AA;
XX
XX Query Match 10.5%; Score 89; DB 23; Length 296;
XX Best Local Similarity 23.1%; Pred. No. 2;
XX Matches 43; Conservative 33; Mismatches 70; Indels 40; Gaps 7;
XX
XX QY 3 VKFPIINKGTGEVSELPKRVYTTIONK-----ENSGTIVSEEDTLP---- 46
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 26 ISQVALGESKNNTTE--PIQLVTVGSDKLLIGLTSHEKPPQLSTEIVLERNFALSHTWK 83
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 47 ----VYKGELEKGYQPDGWEISGFEKGKAGYVINISKQFIKPVFKTIEKKEENKPT 102
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 84 NSGVFSGYKPELIDQLAAGKKGKGLLY-----PPAARKSAKQVNFQLPN 133
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 103 FDVSKKKDN---PQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKST 158
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 134 EDVKAKQDDADSDSEDDSDSDSDSENSGDEEEKVTAESDSEDDSD--DEEDDSSEH 191
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 159 TNNPNK 164
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 192 T--PKK 195
XX
XX RESULT 39
XX ABB64828 standard; Protein; 281 AA.
XX ID ABB64828
XX
XX AC ABB64828;
XX
XX DE 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster polypeptide SEQ ID NO 21276.
XX
XX DE Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX KW Drosophila melanogaster.
XX OS
XX OS Drosophila melanogaster.
XX PN
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX
XX PR 23-MAR-2000; 2000US-191637P.

```

```

PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX PI
XX PI Venter JC, Adams M, Li PMD, Myers EW;
XX MPI; 2001-656860/75.
XX DR N-PSDB; ABL089311.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 21276; 21pp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 281 AA;
XX
XX Query Match 10.4%; Score 88.5; DB 22; Length 281;
XX Best Local Similarity 22.1%; Pred. No. 2.1;
XX Matches 31; Conservative 34; Mismatches 52; Indels 23; Gaps 5;
XX
XX QY 29 NKGMSSTIVSEEDTLPVYKGELEKGYQPDGWEISGFEKGKAGYVI-----NLS 79
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 137 NFEETFMNTTQODLKNLY-SLIDK-----DNEGATSKELGNVIRALGRPNESTA 188
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 80 KDTFIPKPVFKTIEE-KKEENKPTPDVSKKKDNQVNHSQLNE-----SHRKEDLQREH 133
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 189 KEFPCVILIRKHQHDNTKEBELDAPRVFPDKENNGYISTELAAVFMALGKLEDDLEEM 248
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 134 SOKSDSTKDVATVLDKNNI 153
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 249 IREYDLDQDNHINFEESNM 268
XX
XX RESULT 40
XX AAG85008 standard; Protein; 1141 AA.
XX ID AAG85008
XX
XX AC AAG85008;
XX
XX DE 11-SEP-2001 (first entry)
XX
XX DE Shrimp white spot Bacilliform virus (WSBV) protein 99.
XX
XX KW Shrimp white spot Bacilliform virus; MSBV; diagnosis; viral infection;
XX KW antiviral agent; gene expression; antisense construct;
XX KW transgenic viral resistant shrimp.
XX
XX OS
XX OS White spot syndrome virus.
XX PN
XX PN WO200138351-A2.
XX
XX PD 31-MAY-2001.
XX
XX PF 08-NOV-2000; 2000WO-US28888.
XX
XX PR 24-NOV-1999; 99CN-0124717.
XX
XX (PENY-) PE CORP NY.
XX PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
XX (SINO-) SINOGENOMAX CO LTD.

```

XX Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;  
 PI  
 XX WPI; 2001-355877/37.  
 DR N-PSDB; AAH62788.  
 XX

PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus  
 PT (WSBV), useful for producing viral polypeptides that can be used to  
 PT screen for agents that are useful for treating WSBV infection -  
 XX

PS Claim 1; Figure 3; 626pp; English.  
 XX

CC The invention provides the primary nucleotide sequence of the WSBV genome  
 CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and  
 CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences  
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid  
 CC molecules and proteins of the invention are useful for diagnosis and  
 CC monitoring viral infection, in screens for antiviral agents and for  
 CC monitoring viral gene expression or activity during a treatment regimen.  
 CC The nucleic acid molecules are also useful as antisense constructs to  
 CC control viral gene expression in infected cells and tissues and to create  
 CC transgenic viral resistant shrimp.  
 XX

SO Sequence 1141 AA;

Query Match 10.4%; Score 88.5; DB 22; Length 1141;  
 Best Local Similarity 23.3%; Pred. No. 15;  
 Matches 30; Conservative 31; Mismatches 43; Indels 25; Gaps 4;

QY 35 STIVSEEDFILPVYKGELEKGYQFDGMEISGPEG--KRDAGVYINLSKDTFIKPVFKIE 92  
 DB 621 SNVEEBE-----EEQMEEBEEBEVEEREBGSDKDDGDA-----PAQEME 662  
 QY 93 EKKEEENKPTFVSKKQNPQVNHSQLNESHKREDLQREHSQKSDSTQDVTATVLDKNN 152  
 DB 663 EKKEEENQ-----QQPEEESNGNENQEEQQQQQPEREENKMDADSDSDSSSSSSSS 717  
 QY 153 ISSKSTTNN 161  
 DB 718 SSSSSSSSSS 726

Search completed: February 10, 2004, 10:53:56  
 Job time : 38.4525 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 7.62413 Seconds  
(without alignments)  
1011.574 Million cell updates/sec

Title: US-10-067-385-8\_Copy\_610\_773

Perfect score: 848  
Sequence: 1 TTVEFLINKDTGSEVSELPK.....ATVLKNNISSKSTNNPNK 164

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97.5	11.5	1345	1 YH00 YEAST	P38800 saccharomyc
2	97.5	11.5	2468	1 MABP HUMAN	P46821 homo sapien
3	95	11.2	348	1 CVL2 HUMAN	O14093 homo sapien
4	95	11.2	1398	1 TOR2 PLARK	P41001 plasmodium
5	93.5	11.0	649	1 HS70 PYRSA	P37899 pyrenomonas
6	93	11.0	2464	1 MABP MOUSE	P14873 mus musculu
7	92.5	10.9	1888	1 YDR2 SCHPO	O14207 schizosach
8	92	10.8	1202	1 DPMO ASCIM	P22374 ascobolus i
9	92	10.8	1875	1 MPE1 YEAST	O02455 saccharomyc
10	91	10.7	893	1 GYRA STAEF	P54112 staphylococ
11	90.5	10.7	443	1 DACA BACSU	P08750 bacillus su
12	90.5	10.7	1332	1 SPY7 YEAST	P35177 saccharomyc
13	90	10.6	688	1 LIP STAEF	O02510 staphylococ
14	89.5	10.6	433	1 TIG STAEF	O09916 staphylococ
15	87.5	10.3	655	1 SKM1 YEAST	O12469 saccharomyc
16	87.5	10.3	1702	1 TGA2 HAEIN	P43384 haemophilus
17	87	10.3	778	1 YFR8 YEAST	P43610 saccharomyc
18	87	10.3	6632	1 UN89 CAEEL	O01761 caenorhabdi
19	86.5	10.2	406	1 NSB1 MOUSE	O91335 mus musculu
20	86.5	10.2	1006	1 RAT1 YEAST	O02792 saccharomyc
21	86	10.1	443	1 GLNA PYRKO	O08467 pyrococcus
22	86	10.1	914	1 PBPA BACSU	P39793 bacillus su
23	86	10.1	5596	1 MDN1 HUMAN	O9nn22 homo sapien
24	85.5	10.1	752	1 DRS1 YEAST	P32892 saccharomyc
25	85.5	10.1	1007	1 RGAL YEAST	P39083 saccharomyc
26	85.5	10.1	2459	1 MABP RAT	P15055 rattus norv
27	85	10.0	719	1 YBLE SCHPO	O10342 schizosach
28	85	10.0	720	1 IP2 STAEF	O08624 staphylococ
29	84.5	10.0	657	1 Y040 MYCPN	P44963 mycoplasma
30	84.5	10.0	1694	1 TGAO HAEIN	P44963 haemophilus
31	84	9.9	510	1 HG2V HALRO	O02508 haemophilus
32	84	9.9	1251	1 RBR2 PLAVB	O00799 plasmodium
33	83.5	9.8	360	1 FENR_PBA	P10933 pibum sativ

34	83.5	9.8	479	1 U2R1 HUMAN	Q15695 homo sapien
35	83.5	9.8	678	1 GARP PLAVF	P13816 plasmodium
36	83.5	9.8	731	1 HS9A HUMAN	P07900 homo sapien
37	83.5	9.8	943	1 ARS2 DROME	O9v9k7 drosophila
38	83	9.8	258	1 TP31 HUMAN	O75822 homo sapien
39	83	9.8	270	1 TONB HAEIN	P42872 haemophilus
40	83	9.8	715	1 ADSV HUMAN	O27474 homo sapien
41	83	9.8	847	1 DNLI CAEEL	O27474 caenorhabdi
42	83	9.8	3418	1 BRC2 HUMAN	P51587 homo sapien
43	82.5	9.7	439	1 GLNA PYRAB	O9v99 pyrococcus
44	82.5	9.7	472	1 6PGD LACIA	O9cnu6 lactococcus
45	82.5	9.7	1553	1 TP2A CHICK	O42130 gallus galli

## ALIGNMENTS

## RESULT 1

ID	YH00 YEAST	STANDARD	PRT	1345 AA.
AC	P38800	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	Hypothetical 149.7 kDa protein in IRR1-KSP1 intergenic region.			
GN	YHR080C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OX	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
NC	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
RX	MEDLINE=94378003; PubMed=8091229;			
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,			
RA	Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kixeten J.,			
RA	Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,			
RA	Lacaille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,			
RA	Nhan M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,			
RA	Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,			
RA	Vaudin M.;			
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome			
RT	VIII."			
RL	Science 265:2077-2082(1994).			
CC	-I- SIMILARITY: TO YEAST YFL042C.			

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----

EMBL: U10556; AAB68895.1; -.  
PIR: S46817; S46817.  
SGD: S0001122; YHR080C.  
DR: InterPro: IPR004182; GRAM\_dom.  
DR: Pfam: PF02893; GRAM; 1.  
DR: SMART: SM00568; GRAM; 1.  
KW: Hypothetical protein; Transmembrane.  
FT: TRANSMEM 1198 1218 POTENTIAL.  
SQ: SEQUENCE 1345 AA; 149679 MW; 2FDAB94A66564C2 CRC64;

Query Match 11.5%; Score 97.5; DB 1; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 11;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY	49 KGELKGYQFDGWEISGEK-K-DAGYVINSKDTFIKPVFKIIEKKEENKPTFDVSK 107	:     : : :
DB	1109 KGALIEK-----SVEGQKVSVDVWLSELRLDI-----SPAKSKKPKVKVM 1149	:     : : :
QY	108 KKDNPVNHSQLNSHREKEDLQREHSGKSDSTVDVATVLD--KNNISSKSTN---NP 162	:     : : :

Db 1150 SHKHKPHSKVE-----QKSSSRKSDNDKDLITLIDPVQNNFSEIRPMNTLSP 1201

QY 163 NK 164

Db 1202 QK 1203

RESULT 2

MAPB\_HUMAN STANDARD; PRT; 2468 AA.

AC P46821;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Microtubule-associated protein 1B (MAP 1B) [contains: MAP1 light chain LC1].

DE MAP1B.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RX MEDLINE=95104835; PubMed=7806212;

RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;

RT "Cloning of human microtubule-associated protein 1B and the identification of a related gene on chromosome 15.";

RL Genomics 22:273-280(1994).

CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN. PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN STABILIZING MICROTUBULES.

CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE WITH MAP1A AND MAP1B PROTEINS.

CC -1- DOMAIN: Has a highly basic region with many copies of the sequence KKEE and KKEI/V, repeated but not at fixed intervals, which is responsible for the binding of MAP1B to microtubules.

CC -1- PPM: LC1 is coexpressed with MAP1B. It is a polypeptide generated from MAP1B by proteolytic processing. It is free to associate with both MAP1A and MAP1B. It interacts with the amino-terminal region of MAP1B (by similarity).

CC -1- SIMILARITY: TO MAP1A.

CC CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC CC

DR EMBL; L06237; AAA18904.1; -.

DR Genew; HGNC:6836; MAP1B.

DR MIM; 157129; -.

DR GO; GO:0005875; C:microtubule associated complex; TAS.

DR InterPro; IPR000102; MAP1B neurxin.

DR Pfam; PF00414; MAP1B neurxin; 10.

DR PROSITE; PS00230; MAP1B NEURAXIN; 6.

KW Microtubules; Repeat; Phosphorylation.

KW CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.

FT REPEAT 1878 1894 MAP1B 1.

FT REPEAT 1895 1911 MAP1B 2.

FT REPEAT 1912 1928 MAP1B 3.

FT REPEAT 1929 1945 MAP1B 4.

FT REPEAT 1946 1962 MAP1B 5.

FT REPEAT 1963 1979 MAP1B 6.

FT REPEAT 1997 2013 MAP1B 7.

FT REPEAT 2014 2030 MAP1B 8.

FT REPEAT 2031 2047 MAP1B 9.

FT REPEAT 2048 2064 MAP1B 10.

FT DOMAIN 589 790 IYS-RICH (HIGHLY BASIC, CONTAINS MANY KKEE AND KKEI/V REPEATS).

SQ SEQUENCE 2468 AA; 270618 MW; 540839CDDPF09D461 CRC64;

Query Match 11.5%; Score 97.5; DB 1; Length 2468;

Best Local Similarity 24.8%; Pred. No. 21;

Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEPIANKDGEVSELKPHRYVTITONKEMSTIYSEEDFILPYVKGELKGYFDGWEI 63

Db 584 EKVMKKKPKVETKTPSVTEKEVPSKEPS-----PV-KAEVA-----EK 623

QY 64 SEFEGKQAGYINISKOTFIKPVFKIEEKEENKPTDYSKKDNPQVHNSH 123

Db 624 QATDKPRAAKETKTKVTP-----EDKKEKEKPKKVAKEDKTPI---KKEKP 675

QY 124 RKEDLQRE-----EHSQKSDSTKDY 143

Db 676 KKEVYKVEYKEIKKEEKEPKKEV 700

RESULT 3

CYL2\_HUMAN STANDARD; PRT; 348 AA.

AC 014093;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cylindrin II (Multiple-band polypeptide II).

DE CYL2 OR CYL2.

GN Homo sapiens (Human).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RX MEDLINE=95255491; PubMed=7737358;

RA Hess H., Held H., Zimbelmann R., Franke W.W.;

RT "The protein complexity of the cytoskeleton of bovine and human sperm heads: the identification and characterization of cylindrin II.";

RL Exp. Cell Res. 218:174-182(1995).

CC -1- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY BE INVOLVED IN SPERMATID DIFFERENTIATION.

CC -1- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.

CC -1- TISSUE SPECIFICITY: Testis.

CC CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC CC

DR EMBL; Z46788; CAA86752.1; -.

DR PIR; I37271; I37271.

DR Genew; HGNC:2583; CYL2.

DR MIM; 604035; -.

DR GO; GO:0005200; P:structural constituent of cytoskeleton; TAS.

KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.

KW DOMAIN 25 347 31 X 3 AA REPEATS OF K-K-X.

FT DOMAIN 157 240 3 X APPROXIMATE TANDEM REPEATS.

FT REPEAT 157 184 1.

FT REPEAT 185 212 2.

FT REPEAT 213 240 3.

SQ SEQUENCE 348 AA; 39079 MW; D86766599C1809E7 CRC64;

Query Match 11.2%; Score 95; DB 1; Length 348;

Best Local Similarity 30.9%; Pred. No. 3.9;

Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;



```
QY 49 KGELEKGYQPDGMEISGFEGKKDAGVYINLSKDTFIKPFVKKEEKEENKPTF---DV 105
DB 205 ESEBKG---GTEKDSKKKKDS---KKGKDSATIEIQAAYAKADKKEDGKKANKGDE 256
QY 106 SK--KKDNPQVNSQLN-----ESHKEDLOREHSQSDSTDP---VTATVLDKNNI 153
DB 257 SKDAKKDAKEIKKKKKKKKSSSTDSKSDVAKY---SKDATADAKVAKKDTKESXA 313
QY 154 SSK 156
DB 314 DSK 316

RESULT 4
TOP2_PLAFK STANDARD; PRT; 1398 AA.
ID TOP2_PLAFK STANDARD; PRT; 1398 AA.
AC P41001;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2.
OS Plasmidium falciiparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94316496; PubMed=8041616;
RA Cheesman S., McAleese S., Goman M., Johnson D., Horrocks P.,
RA Ridley R.G., Kilbey B.J.;
RT "The gene encoding topoisomerase II from Plasmidium falciiparum.";
RL Nucleic Acids Res. 22:2547-2551(1994).
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -1- SIMILARITY: Belongs to the type II topoisomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X79345; -; NOT_ANNOTATED_CDS.
DR HSBP; P06786; IEGW.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR003957; CBPA_NFYB_topis.
DR InterPro; IPR001241; DNA_topoisomII.
DR InterPro; IPR002205; DNA_topoisomI.
DR Pfam; PF00204; DNA_gyraseb.1.
DR Pfam; PF00521; DNA_topoisomI.1.
DR Pfam; PF02518; HATPase_C.1.
DR PRINTS; PR00615; CCATSUBUNTA.
DR PRINTS; PR00418; TP12FAMILY.
DR PRODOM; PD000742; DNA_topoisomI.1.
DR SMART; SM00387; HATPase_C.1.
DR SMART; SM00433; TOP2C.1.
DR SMART; SM00434; TOP2C.1.
DR PROSITE; PS00177; TOPOISOMERASE_II.1.
DR Isoenzyme; P500177; TOPOISOMERASE II; 1.
FT NP BIND 144 149 ATP (POTENTIAL).
FT ACT_SITE 830 830 DNA CLEAVAGE (BY SIMILARITY).
FT DOMAIN 271 281 POLY-ASN.
FT DOMAIN 308 316 POLY-ASN.
```

```
FT DOMAIN 1089 1093 POLY-LYS.
FT DOMAIN 1327 1234 POLY-LYS.
SQ SEQUENCE 1398 AA, 161029 MW, BAAD7BEE88FE5BE9 CRC64;
Query Match 11.2%; Score 95; DB 1; Length 1398;
Best Local Similarity 23.8%; Pred. No. 17;
Matches 36; Conservative 36; Mismatches 53; Indels 26; Gaps 6;

QY 29 NGKEMSTIVSEEDFIL--PYKGELEKGYQPDGMEISGFEGKKDAGVYINLSKDTFIK 86
DB 1133 NEETIAGITVKDYDLSMPILFSLTEK--VEDLLTQLKEKERLELRNITVETMWLK 1189
QY 87 VFKEIEE-----KKEEENKPTPDVSKKDNQVNSQLNESHKEDLOREHSQ 135
DB 1190 DIEKVEAIEFQRNVELSNRESNK--PKVARKQ-----GSSMKKKKKKLSDEESE 1242
QY 136 --KSDSTKDYATATVLDKNNISKSTTNP 163
DB 1243 GGDTSDSSEPLVNTLNKTKNTKTTTSSNN 1273

RESULT 5
HS70_PYRSA STANDARD; PRT; 649 AA.
ID HS70_PYRSA STANDARD; PRT; 649 AA.
AC P37859;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heat shock 70 kDa protein.
GN HSP70.
OS Pyrenomonas salina.
OC Nucleomorph.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Pyrenomonas.
OX NCBI_TaxID=3034;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94268506; PubMed=8208251;
RA Hofmann C.J.B., Rensing S.A., Hauber M.M., Martin W.F., Mueller S.B.,
RA Couch J., McFadden G.I., Iglot G.L., Water U.-G.;
RT "The smallest known eukaryotic genomes encode a protein gene: towards
RT an understanding of nucleomorph functions.";
RL Mol. Gen. Genet. 243:600-604(1994).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X72621; CAA51197.1; -.
DR PIR: S42488; S42488.
DR HSBP; P08109; ICKR.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70.1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PRODOM; PD000089; Hsp70.1.
DR PROSITE; PS00297; HSP70_1.1.
DR PROSITE; PS00329; HSP70_2.1.
DR PROSITE; PS01036; HSP70_3.1.
KW ATP-binding; Heat shock; Nucleomorph.
SQ SEQUENCE 649 AA; 72079 MW; B627B08FF90C9164 CRC64;
Query Match 11.0%; Score 93.5; DB 1; Length 649;
Best Local Similarity 24.5%; Pred. No. 9.7;
Matches 38; Conservative 26; Mismatches 50; Indels 41; Gaps 7;
```

QY 64 SGFEKGDQDA-----GYVINLS---KDTFLKPVFKIEKKEENKPTDPVSKKQNDPQVN 115  
 DB 538 --LDRKLEKSNENANINIRVRD-----EKKEKIQEDDKSIEKYE----- 582  
 QY 116 HSQNLNHRKEDLQREHSQSDSTYDVTATVLDK 150  
 DB 583 --VLEFETNDELEKEVEKEKELNFPANPLISK 615

RESULT 6  
 MABP\_MOUSE STANDARD; PRT; 2464 AA.  
 ID MAPB\_MOUSE  
 AC P14873;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))  
 GN [Contains: MAP1 light chain LC1].  
 OS MAP1B OR MTAP1B OR MTAP5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RP [1]  
 RP SEQUENCE FROM N.A., AND DOMAIN.  
 RX STRAIN=Swiss Webster; TISSUE=Brain;  
 RX MEDLINE=90094539; PubMed=2480963;  
 RA Noble M., Lewis S.A., Cowan N.J.;  
 RT "The microtubule binding domain of microtubule-associated protein  
 MAP1B contains a repeated sequence motif unrelated to that of MAP2  
 and tau.";  
 RL J. Cell Biol. 109:3367-3376(1989).  
 CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDING OF SUBUNITS  
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
 CC STABILIZING MICROTUBULES.  
 CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
 CC WITH MAP1B AND MAP1B PROTEINS.  
 CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
 CC responsible for the binding of MAP1B to microtubules.  
 CC -1- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED  
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH  
 CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION  
 CC OF MAP1B.  
 CC -1- SIMILARITY: TO MAP1A.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X51396; CAA35761.1; --  
 CC PIR: S07549; QRMSP1.  
 CC MGD: MGI:1306778; Mtap1b.  
 CC DR GO: GO:0016358; P: dendrite morphogenesis; IMP.  
 CC DR GO: GO:0001578; P: microtubule bundling; IMP.  
 CC DR InterPro: IPR000102; MAP1B\_neuraxin.  
 CC DR Pfam: PF00414; MAP1B\_neuraxin; 10.  
 CC DR PROSITE: PS00230; MAP1B\_NEURAXIN; 7.  
 CC DR Microtubules; Repeat; Phosphorylation.  
 CC KW CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.  
 CC FT REPEAT 1874 1890 MAP1B 1.  
 CC FT REPEAT 1891 1907 MAP1B 2.  
 CC FT REPEAT 1908 1924 MAP1B 3.  
 CC FT REPEAT 1925 1941 MAP1B 4.  
 CC FT REPEAT 1942 1958 MAP1B 5.  
 CC FT REPEAT 1959 1975 MAP1B 6.

FT REPEAT 1993 2009 MAP1B 7.  
 FT REPEAT 2026 2010 MAP1B 8.  
 FT REPEAT 2027 2043 MAP1B 9.  
 FT REPEAT 2044 2060 MAP1B 10.  
 FT DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
 KKEE AND KKEI/V REPEATS).  
 SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFFBDA87 CRC64;  
 Query Match 11.0%; Score 93; DB 1; Length 2464;  
 Best Local Similarity 27.8%; Pred No. 45;  
 Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;  
 QY 4 KEFLNKDQGEVSELPKPRVTVTTONGKMSSTVSEDFILPVYKGELEKQYQEDGWEI 63  
 DB 584 EKVLVKKDKPVNTESKP---SVT---EKVEVS---KEEG--SPV-KAEVA-----EK 623  
 QY 64 SGFEKGDQAGYINLSKDTFLKPVFK-KKEEKEENKPTDPVSKKQNDPQVNHSLNS 122  
 DB 624 QATESKP-----KTKDKVVKKEIKTKLEEKKE--KPKKVVKKEDTTPV---KDKK 672  
 QY 123 HKREDLQRE-----EHSQSDSTYDVTATVLDKNNISKSTNNPNK 164  
 DB 673 PKREYKKEIKKEIKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK 728

RESULT 7  
 YD22\_SCHPO STANDARD; PRT; 1888 AA.  
 ID YD22\_SCHPO  
 AC 014207;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C6B12.02c in chromosome I.  
 GN SPACB12.02C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OC NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Frazer A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmons M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Graymoprez B.,  
 RA Meltens I., Vanstreels E., Rieger M., Schaefer M., Mehlert-Auer S.,  
 RA Gabel K., Fuchs M., Fritze C., Holzer E., Moesl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lemach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambut S., Purrelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Useery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: 298531; CAB1064.1; -  
DR PIR: T39009; T39009.  
DR GeneDB: SPAC6B12.02c; -  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 697 717 POTENTIAL.  
FT TRANSMEM 755 775 POTENTIAL.  
FT TRANSMEM 866 886 POTENTIAL.  
FT TRANSMEM 915 935 POTENTIAL.  
FT TRANSMEM 977 997 POTENTIAL.  
FT TRANSMEM 1033 1053 POTENTIAL.  
FT TRANSMEM 1336 1356 POTENTIAL.  
FT TRANSMEM 1645 1665 POTENTIAL.  
FT DOMAIN 1662 1665 POLY-LEU  
SQ SEQUENCE 1888 AA; 217432 MW; 8AD3BBCE32397C29 CRC64;

Query Match 10.9%; Score 92.5; DB 1; Length 1888;  
Best Local Similarity 23.1%; Pred. No. 36;  
Matches 39; Conservative 25; Mismatches 76; Indels 29; Gaps 5;

QY 1 TTVEPIANKDTGEVSELKPHRYVTYIIONGKEMSTIYSE-EDFLIPYKGELEKGYOPD 59  
DB 394 TSENFPQINLVANAVASTIPVYRTTKMKKNPKYVEVEKLPDLILESY----- 442  
DB 60 GWEISGFEGKDGAGVYNLSKDTFIKPVFKIEEKEENKRPDPVSKKKNPOVNHSQL 119  
QY 443 -----GKAPKFLRVFARSSSHIP--KMTRRKQDSKRYFEDKESDQVIDQVLS 492  
DB 120 NESHKEDLOREHSQKSDS-TKDYATATVLDKN-----NISKSTNN 161  
QY 493 DWYSKHELVOQSHSYKKPSDSKSGVGNIFSVNSKSHSNIAKTAANN 541  
DB

RESULT 8  
DPOM\_ASCIM STANDARD; PRT; 1202 AA.  
ID DPOM\_ASCIM STANDARD; PRT; 1202 AA.  
AC P22374;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Probable DNA polymerase (EC 2.7.7.7).  
OS Ascobolus immerus.  
OC Mitochondrion.  
OG Plasmid pA12.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;  
OC Pezizales; Ascombolaceae; Ascobolus.  
OX NCBI\_TaxID=5191;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2/1;  
RX MEDLINE=90066356; PubMed=2573821;  
RA Kempken F., Weinhard F., Esser K.;  
RT "In oranello replication and viral affinity of linear,  
RT extrachromosomal DNA of the ascomycete Ascobolus immerus.";  
RL Mol. Gen. Genet. 218:523-530(1989).  
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + {DNA(N).  
CC -1- MICELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER  
CC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG TO  
CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X15982; CA34106.1; -  
DR PIR: S05362; S05362.  
DR InterPro: IPR006172; DNA\_pol\_B.  
DR InterPro: IPR004868; DNA\_pol\_B\_2.  
DR Pfam: PF03175; DNA\_pol\_B\_2; 1.  
DR SMART: SM00466; POLBc\_1.  
DR PROSITE: PS00116; DNA\_POLYMERASE\_B; 1.  
KW Transferase; DNA-directed DNA polymerase; DNA replication;  
KW DNA-binding; Plasmid; Mitochondrion.  
SQ SEQUENCE 1202 AA; 138279 MW; 51D41FCEBDF2CDE CRC64;

Query Match 10.8%; Score 92; DB 1; Length 1202;  
Best Local Similarity 22.5%; Pred. No. 24;  
Matches 42; Conservative 37; Mismatches 76; Indels 32; Gaps 9;

QY 6 FILNK-----DGEVSELKPHRYVTYIIONGKEMSTI---VSEEDF--ILPYKGELEK-- 54  
DB 311 FVNNAKTFEPFGNVRISIOFGAVT-TLTDKETLITLALFLEREDITHYMSYDEGDIDSK 369  
QY 55 -----GYOPDGWEISGFEGKDGAGVYNLSKDTFIKPVFKIE-----EKKEENK-PTF 103  
DB 370 FPKGSLSPDFKPLKTIETGTYANTYFPYKDIIVAKINKINFGLDLPKTMDSKWPNL 429  
QY 104 DVSKKKNPOVNHSQLNESHKEDLOREHSQKSDSTQVATATVLDKNI-----SSK 156  
DB 430 KLNDDKTSGEIRMTIKKNQSDYDI--IGHMIINDGENVITFNRAVDNIIKIFVTDSM 487  
QY 157 STNNPN 163  
DB 488 GNTNDPN 494

RESULT 9  
MLP1\_YEAST STANDARD; PRT; 1875 AA.  
ID MLP1\_YEAST STANDARD; PRT; 1875 AA.  
AC 002455;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin-like protein MLP1.  
GN MLP1 OR YKR095W OR YKR415.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=93247549; PubMed=8483450;  
RA Koelling R., Nguyen T., Chen B.Y., Botstein D.;  
RT "A new yeast gene with a myosin-like heptad repeat structure.";  
RL Mol. Gen. Genet. 237:359-369(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94205265; PubMed=8154186;  
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,  
RA Remacha M., Jimenez A., del Rey F., Ballester J.P.G., Revuelta J.L.;  
RT "The complete sequence of a 15,820 bp segment of Saccharomyces  
RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three  
RT new open reading frames.";  
RL Yeast 9:1349-1354(1993).  
CC -1- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA  
CC REPAIR.  
CC -1- SIMILARITY: SOME, TO THE TPR ONCOGENE.  
CC -1- CAUTION: REF. 2 MISQUOTES THE GENE NAME AS "MLP1".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@ib-sib.ch).
CC -----
CC DR EMBL; X71992; AAA34783.1; -.
CC DR EMBL; X73541; CAA51948.1; -.
CC DR EMBL; Z28320; CAA82174.1; -.
CC DR PIR; S38173; S38173.
CC DR SGD; S0001803; MBL.
CC DR GO; GO:0005635; C:nuclear membrane; IDA.
CC DR GO; GO:0005654; C:nucleoplasm; IDA.
CC DR GO; GO:0006606; P:protein-nucleus import; IDA.
CC KW Coiled coil; DNA repair.
CC FT DOMAIN 531 487 COILED COIL (POTENTIAL).
CC FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
CC FT CONFLICT 301 301 R -> A (IN REF. 1).
CC SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C906867 CRC64;

Query Match 10.8%; Score 92; DB 1; Length 1875;
Best Local Similarity 25.0%; Pred. No. 39;
Matches 48; Conservative 36; Mismatches 62; Indels 46; Gaps 10;

OY 7 LINKDTGVSEILKPRVTVITQNGKEMSGSTIVSEDF-----ILPVYKGE---LEKGYQF 58
DB 660 LLNK---EIQDLYPSKSDISITLKGKXSRIIAERFLLNSTLDLTLEANDQLRKRPDY 716
OY 59 DGEWISGEFGKDA-----GYVINLSKDTFFIKVFKKIEKK-----KEENK 100
DB 717 ---LQNTILTKODSKTHETLINEYVCECKSLGIVETELNLKEEQQLRVAHEKNLKEQLLNK 772
OY 101 PTFDYSKKKDNQVNHQSOLNESHK-KEPL---QREHSHQKSDSTQDVIA-----TVLDKN 151
DB 773 ---LSPKSDSLRIWVTLQTLQKREDLLEBTRKSCQKIDIEDALSELKGETSQKHQ 828
OY 152 NISSKSTNNPN 163
DB 829 HIKQLSEDDNSV 840

RESULT 10
GYRA_STAEP
ID_GYRA_STAEP STANDARD; PRT; 893 AA.
AC_P54112;
DT_01-OCT-1996 (Rel. 34, Created)
DT_28-FEB-2003 (Rel. 41, Last sequence update)
DT_28-FEB-2003 (Rel. 41, Last annotation update)
DE_DNA gyrase subunit A (EC 5.99.1.13).
CN_GYRA OR SE0005.
OS_Staphylococcus epidermidis.
OC_Bacteriae; Firmicutes; Bacillales; Staphylococcus.
OX_NCBI_TaxID=1282;
[1]
RP_SEQUENCE FROM N.A.
RC_STRAIN=ATCC 12228;
RA_Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RT Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
[2]
RP_SEQUENCE OF 1-94 FROM N.A. AND MDTA8ENBSIS OF SER-84.
CC_MEDLINE=92102204; Pubmed=1662072;
RA_Sreedharan S., Peterson L.R., Fisher L.M.;
RT "Ciprofloxacin resistance in coagulase-positive and -negative
RT staphylococci: role of mutations at serine 84 in the DNA gyrase A
RT protein of Staphylococcus aureus and Staphylococcus epidermidis.";
RL Antimicrob. Agents Chemother. 35:2151-2154(1991).
CC_-1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC_-1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC_-1- SUBUNIT: MADE UP OF TWO
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE

```

```

CC      ENZYME FORMS AN A2B2 TETRAMER.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC      or send an email to license@1sb-sib.ch).
CC
CC -----
DR   EMBL; AE016744; AAC03602.1; -.
DR   EMBL; S72603; AAB20672.1; -.
DR   PIR; A49832; A49832.
DR   HSSP; P09097; 1AB4.
DR   InterPro: IPR002205; DNA_topoisolv.
DR   Pfam; PF03988; DNA_gyrase_C_6.
DR   ProDom; PD000742; DNA_topoisolv; 1.
DR   SMART; SMO0434; TOP4c; 1.
DR   TIGRFAMs; TIGR01063; gyra; 1.
KW   Topoisomerase; Isomerase; DNA-binding; Antibiotic resistance;
KW   Complete proteome.
FT   ACT_SITE 123      123      DNA CLEAVAGE (BY SIMILARITY).
FT   MUTAGEN  84       84       S->F; RESISTANT TO CIPROFLOXACIN.
SQ   SEQUENCE  893 AA; 100113 MW; 2A6A7CD345A526CE CRC64;
Query Match      10.7%; Score 91; DB 1; Length 893;
Best Local Similarity 24.9%; Pred. No. 21;
Matches 43; Conservative 31; Mismatches 75; Indels 24; Gaps 8;
QY      1  TTVEKFLINDGTGSESLKPRVTVTLQNGKENS-STIVSEEDFLPVYKGELEKGYQFD 59
Db      727  TPSEYSLNMRGKGKIR-----TATITRRNGNIYCIITVGEEDLMVYTNAGVI---IRLD 779
QY      60  GWEISGEGKKDAGY-VINLSKDTFIKPVFKLEKKEEN-----KPIFDVSKKK 109
Db      780  VHDISQ-NGRAAQCVRLMKLGDQGFVSTAVKVNEDNEENADEAQQSTTTETADVEVV 838
QY      110 DNPQVNHSQLNESRKEDLQRE-HSQKSDSTQDVATVLDKNNISKSTNN 161
Db      839 D----DQTPNVAITTEGDAMESEVSEPNDRIDIRDQFMDRVNEDIASDN 887

RESULT 11
DACA_BACSU      STANDARD;      PRT;      443 AA.
AC      P08750;
DT      01-AUG-1988 (Rel. 08, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      D-alanyl-D-alanine carboxypeptidase precursor (EC 3.4.16.4) (DD-
DE      peptidase) (DD-carboxypeptidase) (Cpase) (PBPs).
GN      DACA.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX      NCBI_TaxID=1423;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168;
RA      MEDLINE=96051385; PubMed=7584024;
RA      Ogasawara N., Nakai S., Yoshikawa H.;
RA      "Systematic sequencing of the 180 kilobase region of the Bacillus
RA      subtilis chromosome containing the replication origin.";
RA      DNA Res. 1:1-14(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168;
RA      MEDLINE=98044033; PubMed=9384377;
RA      Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA      Azevedo V., Bertero M.G., Bessieres P., Bojotin A., Borchert S.,
RA      Bortis R., Boutsier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA      Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA      Choi S.K., Codani J.J., Connerton I.F., Cummings S.D., Daniel R.A.,
RA      Denizot F., Devine K.W., Dusterhoft A., Ehlich S.D., Emerson P.T.,

```

RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz S.Y., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.U.D., Grandi G.,  
 RA Guisepi G., Guy B.U., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Ilaya M., Jones L.,  
 RA Joris B., Karmata D., Kasahara Y., Klaerr-Ilanchand M., Klein C.,  
 RA Kobayashi Y., Koetter P., Konigstein G., Krogsh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigic C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Portetelle S., Prescott A.M.,  
 RA Precourt E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi Y., Sekowska A., Seror S.J., Serier P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi B., Takagi T., Takahashi H., Takemori K.,  
 RA Takuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosoato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitznegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.,  
 RA "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RT Nature 390:249-256(1997).  
 RL [3]  
 RN SEQUENCE OF 32-102.  
 RP MEDLINE=80182289; PubMed=6768745;  
 RX Waxman D.J., Strominger J.L.;  
 RA "Sequence of active site peptides from the penicillin-sensitive D-  
 RT alanine carboxypeptidase of *Bacillus subtilis*. Mechanism of  
 RT penicillin action and sequence homology to beta-lactamases.";  
 RL J. Biol. Chem. 255:3964-3976(1980).  
 RN [4]  
 RN SEQUENCE OF 103-443 FROM N.A.  
 RP MEDLINE=86250602; PubMed=3087956;  
 RX Todd J.A., Roberts A.N., Johnstone K., Pigott P.J., Winter G.,  
 RA Ellier D.J.;  
 RT "Reduced heat resistance of mutant spores after cloning and  
 RT mutagenesis of the *Bacillus subtilis* gene encoding penicillin-binding  
 RT protein 5.";  
 RL J. Bacteriol. 167:257-264(1986).  
 RN [5]  
 RN SEQUENCE OF 414-443.  
 RP MEDLINE=81117303; PubMed=6780559;  
 RX Waxman D.J., Strominger J.L.;  
 RA "Primary structure of the COOH-terminal membranous segment of a  
 RT penicillin-sensitive enzyme purified from two *Bacilli*.";  
 RL J. Biol. Chem. 256:2067-2077(1981).  
 RN [1]  
 RN FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE  
 CC CELL WALL PRECURSORS.  
 CC [1]  
 CC CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O = 2 D-alanine.  
 CC [1]  
 CC PATHWAY: Peptidoglycan synthesis; final stages.  
 CC [1]  
 CC SUBCELLULAR LOCATION: Membrane-associated.  
 CC [1]  
 CC SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11, ALSO KNOWN AS THE  
 CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.  
 CC [1]  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC [1]  
 CC EMBL: D26185; BAA05246.1; -  
 CC EMBL: 293104; CAB11786.1; -  
 CC EMBL: M13766; AAA22375.1; -  
 CC PIR: S66040; S66040.  
 CC MEROPS: S11.001; -  
 CC Subtilisin; BG10074; dact.  
 CC InterPro: IPR001967; Ala/AlaCbp1ase1.  
 CC Pfam: PF00768; Peptidase\_S11; 1.

DR PRINTS; PR00725; DADACBP1ASE1.  
 KW Hydrolyase; Carboxypeptidase; Peptidoglycan synthesis; Cell wall;  
 KM Membrane; Signal; Complete proteome.  
 FT SIGNAL 1 31  
 FT CHAIN 32 443 D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.  
 FT ACT SITE 67 67 ACYLATED BY PENICILLIN.  
 FT CONFLICT 100 100 E -> Q (IN REF. 3).  
 FT CONFLICT 227 227 E -> Q (IN REF. 4).  
 SQ SEQUENCE 443 AA; 48636 MW; DA6CSB0307D0C117 CMC64;  
 Query Match 10.7%; Score 90.5; DB 1; Length 443;  
 Best Local Similarity 24.1%; Pred. No. 11;  
 Matches 32; Conservative 25; Mismatches 49; Indels 27; Gaps 4;  
 QY 6 FLTKNDTGEVSELRPHRTVTITONGKMSSTIVSEEDPILRVYKGELEKGFQDGMELSG 65  
 DB 310 FSMKEIYAGGVQVGHK-TISVDKKEKEVGIVTKKASLPVKNGE-EKNYKAK----- 361  
 QY 66 FEGKADAGVINTLSKDTPIKVFVKKEKEEENKPTDVSKKDNPQVNHSHLSHRK 125  
 DB 362 -----VTLNKDNITAVKKGTKVKG-----LTAETGDEKQVGFPLNSDLAG 402  
 QY 126 EDLQREHSQKSD 138  
 DB 403 VDLVTKNVKEKAN 415  
 RESULT 12  
 SPT7 YEAST STANDARD; PRT; 1332 AA.  
 ID SPT7 YEAST  
 AC P35177;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Transcriptional activator SPT7.  
 GN SPT7 OR YBR081C OR YBR0739.  
 OS *Saccharomyces cerevisiae* (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=S288C;  
 RC MEDLINE=95229044; PubMed=7713415;  
 RX Gansheroff L.J., Dollard C., Tan P., Winston F.;  
 RA "The *Saccharomyces cerevisiae* SPT7 gene encodes a very acidic protein  
 RT important for transcription in vivo.";  
 RL Genetics 139:523-536(1995).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=S288C;  
 RC MEDLINE=95076715; PubMed=7985423;  
 RX van der Aart Q.J.M., Barthe C., Doignon F., Aigle M., Crouzet M.,  
 RA Steensma H.Y.;  
 RT "Sequence analysis of a 31 kb DNA fragment from the right arm of  
 RT *Saccharomyces cerevisiae* chromosome II.";  
 RL Yeast 10:959-964(1994).  
 RN [3]  
 RN SEQUENCE OF 1-835 FROM N.A.  
 RP STRAIN=S288C;  
 RC Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,  
 RA Visiers S.;  
 RT Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN SEQUENCE OF 463-523 FROM N.A.  
 RP MEDLINE=92285152; PubMed=1350857;  
 RX Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J.,  
 RA David I.B.;  
 RT "The bromodomain: a conserved sequence found in human, *Drosophila* and  
 RT yeast proteins.";  
 RL Nucleic Acids Res. 20:2603-2603(1992).  
 CC [1]  
 CC FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY  
 CC OTHER GENES.

```

CC -1- CATALYTIC ACTIVITY: Triacylglycerol + a
CC fatty acid anion.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS
CC NEGATIVELY REGULATED BY BACTERIOPHAGE LYSGENIZATION (LIPASE
CC CONVERSION).
CC -1- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M95577; AAA19729.1; -.
CC DR EMBL; AE016744; AAC03878.1; -.
CC DR PIR; A47705; A47705.
CC DR InterPro; IPR005877; Gpos_Y5IRK.
CC DR InterPro; IPR000734; Lipase.
CC DR InterPro; IPR000379; Ser_estrs_site.
CC DR Pfam; PF04650; Y5IRK_signal; 1.
CC DR TIGRfams; TIGR01168; Y5IRK_signal; 1.
CC DR PROSITE; PS00120; LIPASE_SER; 1.
CC KW Hydrolase; Lipid degradation; Zymogen; Signal; Complete proteome.
CC FT SIGNAL 1 35
CC PROPEP 36 302
CC CHAIN 303 688
CC FT ACT_SITE 418 418
CC FT ACT_SITE 648 648
CC FT CONFLICT 96 96
CC FT CONFLICT 120 120
CC SQ SEQUENCE 688 AA; 77343 MW; 6C95DB3A78AF86F6 CRC64;
CC -----
Query March 10.6%; Score 90; DB 1; Length 688;
Best local Similarity 24.4%; Pred. No. 19;
Matches 39; Conservative 30; Mismatches 65; Indels 26; Gaps 7;
QY 9 NKDTGEVSEIKPHRYVTIIONGKEMSTIIVSEEDFILPYTKGELEKGYQPDGWEISGFE-67
DB 68 NKNVNEKSNVNV-----SITENESLIHNETPKNEWDI-----QQQKDSQNDNKSSESVEEQ 115
QY 68 GKQDAGVIVLKSOTFIKVPFKIE-EKKEENKPTFDVSKKKDNQPV--NHSQLNESHK 124
DB 116 NKEAEVAVQNHSE--KPOQEOVELKHA SENNQTHLSKRAQSNEDVKTGPSOLDNTAA 172
QY 125 KEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPNK 164
DB 173 KQEDSQKENLSKODTQSSKTTDL-----RATQNGSK 205
RESULT 14
TIG_STAM STANDARD; PRT; 433 AA.
AC Q99T16;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Trigger factor (TF).
GN TIG OR SAV1675 OR SAL1499 OR MM1619.
OS Staphylococcus aureus (strain M50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MM2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxId=158878, 158879, 196620;
[1]
SEQUENCE FROM N.A.
RP
RC STRAIN=M50 / ATCC 700699, and N315;
RX MEDLINE=21311952; Pubmed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hasegawa A.,

```







DR SGP: S0001934; YFRO38W.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR000330; SNF2\_N.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00176; SNF2\_N; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELICC; 1.  
DR Hypothetical protein; Nuclear protein; DNA-binding; Helicase;  
KM ATP-binding.  
FT NP BIND 247 254 ATP (POTENTIAL).  
FT SITE 352 355 DECH BOX.  
SQ SEQUENCE 778 AA; 88730 MW; 356C0857B5EAD84 CRC64;  
  
Query Match 10.3%; Score 87; DB 1; Length 778;  
Best Local Similarity 24.2%; Pred. No. 35;  
Matches 44; Conservative 26; Mismatches 70; Indels 42; Gaps 6;  
  
QY 4 KEFLNKDGVSELSKPRVVTITQNGKEMST-----IVSEDFILPYKGELEKG 55  
DB 35 REKEVNDLTLADISDSDDLSDSDKRGKNDTAPITLQDDVHSDEDI----- 82  
QY 56 YQPDGMEISGFEKGDAGVNIKSDTFIKPVFKIEKEEENKPTDVSKKDNPQVN 115  
DB 83 -QLDSEDSDTREAVG-AQVVDKLAQDT--KSFQKSLDELSDMDTKIVSLKAKKINFEVR 138  
QY 116 HSQLES-----HRKEDLQR-----EEHSQKSDTKQVATVLDKNNISSKS 157  
DB 139 OSQVYSIIADTLTLRSNEVANANTKDNNSDDEHSSKKRTKKKSLTDFKQKKNED 198  
QY 158 TT 159  
DB 199 TT 200  
  
UN89\_CABEL STANDARD; PRT; 6632 AA.  
ID UN89\_CABEL  
AC 001761; Q17362;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).  
GN UNC-89 OR C09D1.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_Taxid=6239;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
RC STRAIN-Bristol N2;  
RX MEDLINE=96180278; PubMed=8603916;  
RA Benjan G.M., Tinley T.L., Tang X., Borodovsky M.;  
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line  
assembly, encodes a giant modular protein composed of Ig and signal  
transduction domains.";  
RT J. Cell Biol. 132:835-848 (1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RA Du Z., Le T.T., Wilson R.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Structural component of the muscle M-line. Myofibril  
lattice assembly begins with positional cues laid down in the  
basement membrane and muscle cell membrane. UNC-89 responds to  
these signals, localizes, and then participates in assembling an  
M-line.  
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.  
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.

CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.  
CC -!- SIMILARITY: Contains 1 PH domain.  
CC -!- SIMILARITY: Contains 5 RCDSD domains.  
CC -!- SIMILARITY: Contains 1 SH3 domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb.ch/announce/>  
or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC -----  
DR EMBL; U33058; AAB00542.1; -.  
DR EMBL; AF003131; AAB54132.2; -.  
DR PDB; 1PHO; 20-DEC-00.  
DR WormPep; C09D1.1; CE30426.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR007110; Ig\_1-like.  
DR InterPro; IPR003598; Ig\_C2.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR01849; PH.  
DR InterPro; IPR007850; RCDSD.  
DR InterPro; IPR000219; RHOGEF.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00041; Im3; 1.  
DR Pfam; PF00047; Ig; 47.  
DR Pfam; PF00169; PH; 1.  
DR Pfam; PF05177; RCDSD; 5.  
DR Pfam; PF00621; RHOGEF; 1.  
DR Pfam; PF00018; SH3; 1.  
DR SMART; SM00408; IGC2; 23.  
DR SMART; SM00325; RHOGEF; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS00010; DH\_2; 1.  
DR PROSITE; PS00835; IG\_LIKE; 49.  
DR PROSITE; PS00003; PH\_DOMAIN; 1.  
DR PROSITE; PS00002; SH3; 1.  
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;  
KW 3D-structure.  
FT DOMAIN 63 127 SH3.  
FT DOMAIN 152 330 DH.  
FT DOMAIN 342 498 PH.  
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.  
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.  
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.  
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.  
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.  
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.  
FT DOMAIN 1272 1315 THR-RICH.  
FT DOMAIN 1375 1475 RCDSD 1.  
FT DOMAIN 1479 1585 RCDSD 2.  
FT DOMAIN 1597 1695 RCDSD 3.  
FT DOMAIN 1700 1799 RCDSD 4.  
FT DOMAIN 1800 1860 RCDSD 5.  
FT DOMAIN 1882 2067 IG-LIKE C2-TYPE 7.  
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.  
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.  
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.  
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.  
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.  
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.  
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.  
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.  
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.  
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.  
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.  
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.  
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.  
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.  
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.  
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.

```

FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 6171 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT DISULFID 2137 2137 AKA -> P (IN REF. 1).
FT CONFICT 2245 2247 A -> P (IN REF. 1).
FT CONFICT 2258 2258 E -> G (IN REF. 1).
FT CONFICT 2284 2284 W -> I (IN REF. 1).
FT CONFICT 2297 2297 A -> G (IN REF. 1).
FT CONFICT 3531 3531 DAGEY -> RRRRI (IN REF. 1).
FT CONFICT 3884 3888 A -> V (IN REF. 1).
FT CONFICT 3929 3929 A -> P (IN REF. 1).
FT CONFICT 5134 5134 T -> S (IN REF. 1).
FT CONFICT 5145 5145 G -> A (IN REF. 1).
FT CONFICT 5185 5185 K -> N (IN REF. 1).
FT CONFICT 5199 5199 L -> F (IN REF. 1).
FT CONFICT 5202 5202 F -> L (IN REF. 1).
FT CONFICT 5213 5213 A -> G (IN REF. 1).
FT CONFICT 6178 6178 K -> E (IN REF. 1).
FT CONFICT 6268 6268
SQ SEQUENCE 6632 AA; 731665 MW; 262D3BD62960E89 CRC64;

```

```

Db 1242 ESVVEKODLSSS----EYOKIEIAQVKEASEATTITMETSLSTKT 1287
RESULT 19
ID NSB1_MOUSE STANDARD; PRT; 406 AA.
AC Q9JL35; O88832; Q8VC71; Q9CWM1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 26-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Nucleosome binding protein 1 (Nucleosome binding protein 45) (NBP-45) (GARP45 protein).
GN NSB1 OR GARP45.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=20158948; PubMed=10692437;
RA Shirakawa H., Landsman D., Postnikov Y.V., Bustin M.;
RT "NBP-45, a novel nucleosomal binding protein with a tissue-specific
and developmentally regulated expression."
RL J. Biol. Chem. 275:6368-6374(2000).
[2]
RP SEQUENCE FROM N.A.
RA Onoda G., Suzuki N., Saito H., Honda T., Sato H., Kuwano R.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBD databases.
[3]
RP TISSUE-COLON;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schefer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.D., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,
RA Villalón D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[4]
RP SEQUENCE OF 7-195 FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=hippocampus;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai T., Shinagawa A., Shibata K., Yoshino M., Ishii Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamata I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glass C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,
RA Schiraldi L.M., Staudt R., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balasubramanian R., Barsh G.,
RA Blake J., Bojelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.D., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofman M., Hume D.A., Kamaya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima U., Mazzarelli I., Komabaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto K.,
RA Sasaki H., Sato K., Schenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,

```

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
CC -1- FUNCTION: Binds specifically to nuclear core particles and acts as  
CC a transcriptional activator.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: Widely expressed with highest levels in  
CC submaxillary gland, thymus, kidney and liver and lowest levels in  
CC brain, lung, pancreas and eye.  
CC -1- DEVELOPMENTAL STAGE: Highest levels are found in 7-day-old  
CC embryos. Levels in the 7-day-old embryo are 4-fold higher than in  
CC the adult and almost 10-fold higher than in later embryonic  
CC stages.  
CC -1- SIMILARITY: BELONGS TO THE HMG-14/HMG-17 PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL, AF213454; AAF30179.1; -;  
CC EMBL, AB018374; BAA33783.2; -;  
CC EMBL, BC021626; AAB21626.1; -;  
CC EMBL, AK013748; BAB28982.1; -;  
CC MGI, MGI:1355295; Nbdpl.  
CC GO, GO:0005654; C:nucleoplasm; IDA.  
CC GO, GO:0003682; P:chromatin binding activity; IDA.  
CC GO, GO:0006356; P:regulation of transcription from Pol I prom. .; IDA.  
CC InterPro: IPR000079; HMG\_14\_17.  
CC DR Pfam, PF01101; HMG14\_17; 1.  
CC DR SMART, SM00527; HMG17; 1.  
CC DR PROSITE, PS00355; HMG14\_17; FALSE NEG.  
CC DR Transcription regulation, Activator; DNA-binding; Nuclear protein.  
CC FT CONFLICT 74 74 M -> V (IN REF. 1).  
CC FT CONFLICT 390 390 N -> H (IN REF. 2).  
CC SQ SEQUENCE 406 AA; 45344 MW; 5944305613EC9679 CRC64;  
CC -----  
CC Query Match 10.2%; Score 86.5; DB 1; Length 406;  
CC Best Local Similarity 28.0%; Pred. No. 19;  
CC Matches 44; Conservative 16; Mismatches 52; Indels 45; Gaps 8;  
CC -----  
CC QY 9 NKTGEVSELKPHRVTVITQNGK-----ENSSITVSEEDFLPYKGELEKGYQFDGW 61  
CC DB 141 HKDTGE-----EVEDGKIEBEGLEKPGTAKSBD--AEVSKDEEKGDNKNG- 185  
CC QY 62 EISGEGKKDAGVYNLSKD-----TFIKPYPKKIEEK---KEENKPTDVSCK 108  
CC DB 186 -----EDGEKEDEKEEEDKDEGDTGTEKEVKEONKEAEEDDKCKEENK---EVGKE 237  
CC QY 109 KDNPOVNSQLNESHKEDLQRE--HSQKSDSTKDV 143  
CC DB 238 GQPEEDGKEDLHEVKGEDLHEEDGKEGQPEEDGKEI 274  
CC -----  
CC RESULT 20  
CC RAT1 YEAST  
CC ID RAT1 YEAST STANDARD; PRT; 1006 AA.  
CC AC 002792;  
CC DT 01-OCT-1993 (Rel. 27, Created)  
CC DT 01-OCT-1993 (Rel. 27, Last sequence update)  
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)  
CC DE Ribonucleic acid trafficking protein 1 (5'-3' exoribonuclease)  
CC (EC 3.1.11.-) (P116)  
CC GN RAT1 OR HKR1 OR TRP1 OR YOR048C.  
CC OS Saccharomyces cerevisiae (Baker's yeast).  
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
CC CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
CC OX NCBI\_TaxId=4932;  
CC RN [1]

RP SEQUENCE FROM N.A. PubMed=1628825;  
RX MEDLINE=92331925; PubMed=1628825;  
RA Ambler D.C.; Goldstein A.L.; Cole C.N.;  
RT "Isolation and characterization of RAT1: an essential gene of  
RT Saccharomyces cerevisiae required for the efficient nucleocytoplasmic  
RT trafficking of mRNA."  
RL Genes Dev. 6:1173-1189(1992).  
CC (2)  
CC SEQUENCE FROM N.A. AND EXORIBONUCLEASE ACTIVITY.  
CC MEDLINE=93109318; PubMed=8417335;  
CC Kenna M., Stevens A., McCammon M., Douglas M.G.;  
CC "An essential yeast gene with homology to the exonuclease-encoding  
CC XRN1/XEM1 gene also encodes a protein with exoribonuclease  
CC activity."  
CC Mol. Cell. Biol. 13:341-350(1993).  
CC (3)  
CC SEQUENCE FROM N.A.  
CC MEDLINE=93268292; PubMed=8497260;  
CC Aldrich T.L., di Segni G., McConaughy B.L., Keen N.J., Whelan S.,  
CC Hall B.D.;  
CC "Structure of the yeast TRP1 protein: dependence of transcription  
CC activation on the DNA context of the target gene."  
CC Mol. Cell. Biol. 13:3434-3444(1993).  
CC (4)  
CC SEQUENCE FROM N.A.  
CC Landt O., Hiesel R., Unseld M.;  
CC Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.  
CC (5)  
CC SEQUENCE FROM N.A.  
CC Bohn C., Bolotin-Fukuhara M., Daignan-Fornier B., Dang D.V.,  
CC Valens M.;  
CC Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.  
CC (6)  
CC FUNCTION: May function in the processing and/or trafficking of  
CC nuclear mRNA. May be involved in general transcription as well.  
CC Possesses 5'->3' exoribonuclease activity degrading poly(A) to  
CC mainly 5'-AMP.  
CC -1- COFACTOR: Requires magnesium.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: Belongs to the 5'-3' exonuclease family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL, S61567; AAB26818.1; -;  
CC EMBL, M95626; AAA34860.1; -;  
CC EMBL, L06011; AAA16950.1; -;  
CC EMBL, Z11746; -; NOT ANNOTATED CDS.  
CC EMBL, Z74956; CAA9240.1; -;  
CC PIR, S20126; S20126.  
CC SGD, S0005574; RAT1.  
CC GO, GO:0005634; C:nucleus; IDA.  
CC GO, GO:0004534; F:5'-3' exoribonuclease activity; IDA.  
CC GO, GO:0006365; P:35S primary transcript processing; IMP.  
CC GO, GO:0006396; P:RNA processing; IMP.  
CC InterPro: IPR004859; Put\_53exo.  
CC Pfam, PF01159; XRN\_N\_1.  
CC KM Nuclear protein; Hydroxylase; Nuclease; Exonuclease; Repeat.  
CC FT DOMAIN 525 528 POLY-GU.  
CC FT DOMAIN 793 797 POLY-ASN.  
CC FT DOMAIN 955 999 CONTAINS 2 X SRVD, 2 X NNNY, AND 2 X  
CC MUTAGEN 683 683 YSGN REPEATS.  
CC Y->H. IN ALLELE TRP1-1; ACTIVATES  
CC TRANSCRIPTION OF THE PROMOTER-DEFECTIVE  
CC YEAST SUP4 TRNA(TYR) ALLELE SUP4A53T61.  
CC FT FT  
CC SQ SEQUENCE 1006 AA; 115933 MW; 5DD5B0245F3E12A CRC64;  
CC Query Match 10.2%; Score 86.5; DB 1; Length 1006;  
CC Best Local Similarity 19.0%; Pred. No. 50;

Matches 32; Conservative 40; Mismatches 47; Indels 49; Gaps 8;

QY 27 IONKMSSTIVSEEDFILPVY--KGELEKGYFPGDWEISGPEGKADAGVILNLDSTFI 84  
 Db 415 MSKGDHRHPFVATEQ--LQMYDTQGNLAKG--SWNLITSD-----MYRLKKEML 460  
 QY 85 -----KRVFKKIEKEEKKPFVDVSKKKNDPQVNSHQ-----LNSHRK 125  
 Db 461 ANEGNEBAIAKVAQSDKNNELMKDISKEEIDAVSKANKTNPNLAEWKOKITINKKHL 520

QY 126 EDLOREEH-----SOKSDSTKDVATVLDK-----NNISKSST 159  
 Db 521 EKONEEBELAKSKVYKTEKASECDLDEIKDEIVADVDRNSEET 568

RESULT 21  
 GLNA PYRKO STANDARD; PRT; 443 AA.  
 AC 008467;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
 GN GLNA.  
 OS Pyrococcus kodakarensis.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Thermococcus.  
 OX NCBI\_TaxID=69014;  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN=KOD1.  
 RX MEDLINE=97316461; PubMed=9172372;  
 RA Abdul Rahman R.N.Z., Jongsareejit B., Fujiwara S., Imanaka T.;  
 RT "Characterization of recombinant glutamine synthetase from the  
 RT hyperthermophilic archaeon Pyrococcus sp. strain KOD1.";  
 RL Appl. Microbiol. 63:2472-2476(1997).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
 CC L-glutamine.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; D86222; BAA20530.1; -.  
 DR HSSP; P06201; IUGR.  
 DR InterPro; IPR001691; GUN synth.  
 DR InterPro; IPR004809; GlnA.  
 DR InterPro; IPR001637; GlnA adenyltn.  
 DR Pfam; PF00120; gln-synt; 1.  
 DR Pfam; PF03951; gln-synt; 1.  
 DR ProDom; PD001057; Gln\_synt\_C; 1.  
 DR TIGRFAMs; TIGR00653; GlnA\_1.  
 DR PROSITE; PS00180; GlnA\_1; 1.  
 DR PROSITE; PS00181; GlnA\_ATP; 1.  
 KM ligase.  
 FT BINDING 362 362 AMP (UNDER CONDITIONS OF ABUNDANT  
 FT GLUTAMINE) (BT SIMILARITY).  
 FT SEQUENCE 443 AA; 50259 MW; 9426DCCEFEF18168 CRC64;

Query Match 10.1%; Score 86; DB 1; Length 443;  
 Best Local Similarity 17.8%; Pred. No. 22;  
 Matches 31; Conservative 23; Mismatches 46; Indels 72; Gaps 4;

QY 48 YKGELEKGYFPGDWEISGPEGKADAGVILNLDSTFI----- 84  
 Db 42 YEEAVDEGVSPDSSISGFGIDSDLIFRADSTYALIEIWEGIGRYGYTKGDEPYQA 101

QY 85 -KRVFKKIEKEEKEN-----KPTFVSKKCD----- 110  
 Db 102 DRGLIKVTLERLEKEGKAHNGPEPEFYIFKKNQWELHLPDSGCVFDVLGDKAREIR 161  
 QY 111 -----NPOVNSQLNESHKRDLOREHSQSD--STKDYATATVLD 149  
 Db 162 REIALYMPYLGLKPEVLHHEVGAKQHEIDFREDALRTADNIVSPKRVKAVAE 215

RESULT 22  
 BBPA BACSU STANDARD; PRT; 914 AA.  
 ID BBPA BACSU  
 AC P39793;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Penicillin-binding protein 1A/1B (PBP1) (includes: Penicillin-  
 DE insensitive transglycosylase (EC 2.4.2.-) (peptidoglycan Tgasee);  
 DE penicillin-sensitive transpeptidase (EC 3.4.-.-) (DD-transpeptidase)].  
 GN PONA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 499-515.  
 RC STRAIN=168;  
 RX MEDLINE=95113769; PubMed=7814321;  
 RA Popham D.L., Setlow P.;  
 RT "Cloning, nucleotide sequence, and mutagenesis of the Bacillus  
 RT subtilis pona operon, which codes for penicillin-binding protein  
 RT (PBP) 1 and a PBP-related factor.";  
 RL J. Bacteriol. 177:326-335(1995).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=168 / Marburg;  
 RX MEDLINE=96349105; PubMed=8760912;  
 RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,  
 RA Serror P.;  
 RT "Sequence analysis of the Bacillus subtilis chromosome region between  
 RT the *se* and *kds* loci cloned in a yeast artificial chromosome.";  
 RL Microbiology 142:2005-2016(1996).  
 RL [3]  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunes T., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Betteiro M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Eutlan K.D., Errington J., Fadre C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Goldlighty E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blandhard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Puig J., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivoita C., Roche E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiuchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccoci E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Teperstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandendol M., Vannier F., Vasseroiti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,

RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT subtilis".  
 RL Nature 390:249-256 (1997).  
 RN [4]  
 RP GROWTH REQUIREMENTS.  
 RC STRAIN=168;  
 RX MEDLINE=98389671; PubMed=9721295;  
 RA Murray T., Popham D.L., Setlow P.;  
 RT "Bacillus subtilis cells lacking penicillin-binding protein 1 require  
 RL increased levels of divalent cations for growth.";  
 RL J. Bacteriol. 180:4555-4563 (1998).  
 RN [5]  
 RP SUBCELLULAR LOCATION.  
 RC STRAIN=168;  
 RX MEDLINE=99255546; PubMed=10322023;  
 RA Pedersen L.B., Angert E.R., Setlow P.;  
 RT "Septal localization of penicillin-binding protein 1 in *Bacillus*  
 RL subtilis".  
 RL J. Bacteriol. 181:3201-3211 (1999).  
 CC -1- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED  
 CC PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A  
 CC PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN  
 CC (FORMATION OF LINEAR GLYCAN STRANDS) AND A PENICILLIN-SENSITIVE  
 CC TRANSPEPTIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE  
 CC SUBUNITS) (BY SIMILARITY).  
 CC -1- PATHWAY: Peptidoglycan synthesis; final stages.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. PROBABLY FOUND ALL  
 CC OVER THE WHOLE CELL AT LOW CONCENTRATIONS. ALSO LOCALIZES TO THE  
 CC DEVELOPMENTAL STAGE: EXPRESSION IS CONSTANT DURING GROWTH,  
 CC DECREASES DURING SPOULATION AND IS INDUCED APPROXIMATELY 15 MIN  
 CC INTO SPORE GERMINATION.  
 CC -1- PTM: THE PRODUCT EXPRESSED FROM THE TRANSLATION OF THE PONA GENE  
 CC APPEARS AS TWO BANDS ON A GEL (1A AND 1B), BUT THE SPECIFIC AMINO  
 CC ACID SEQUENCE OF EACH PROTEIN IS UNKNOWN.  
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -1- MISCELLANEOUS: CELLS LACKING THE PROTEIN REQUIRE INCREASED LEVELS  
 CC OF MG(2+) OR CA(2+) FOR GROWTH AND GERMINATION. APPROXIMATELY 50%  
 CC OF CELLS WITHOUT THE PROTEIN CONTAIN ABNORMAL FT52 RINGS,  
 CC SUGGESTING IT IS INVOLVED IN SEPTUM SYNTHESIS; INCREASED LEVELS OF  
 CC MG(2+) OR CA(2+) ONLY PARTIALLY ELIMINATE THE SEPTATION DEFECTS.  
 CC -1- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE  
 CC TRANSGLYCOSYLASE FAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE  
 CC TRANSPEPTIDASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U11883; AAA64947.1; -;  
 CC DR EMBL; L47838; AAB38459.1; -;  
 CC DR EMBL; Z99115; CAB14148.1; -;  
 CC DR PIR; I40529; I40529.  
 CC DR Subtilast; BG10954; PONA.  
 CC DR InterPro; IPR003961; FN\_III.  
 CC DR InterPro; IPR001264; Glyco\_trans\_51.  
 CC DR InterPro; IPR001460; Transpeptidase.  
 CC DR Pfam; PF00041; fn3; 1.  
 CC DR Pfam; PF00912; Transglycosyl; 1.  
 CC DR Pfam; PF00905; Transpeptidase; 1.  
 CC DR ProDom; PD001895; Glyco\_trans\_51; 1.  
 CC DR SMART; SM00060; FN3; 1.  
 CC DR Peptidoglycan synthase; Cell wall; Transferase; Glycosyltransferase;  
 CC Hydrolase; Multifunctional enzyme; Transmembrane; Signal-anchor;  
 CC Antibiotic resistance; Complete proteome.  
 CC FT DOMAIN 1 37 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 38 58 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)

FT FT (POTENTIAL).  
 FT DOMAIN 59 914 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 77 246 TRANSGLYCOSYLASE.  
 FT DOMAIN 329 662 TRANSPEPTIDASE.  
 FT DOMAIN 706 784 FIBRONECTIN TYPE-III.  
 FT ACT SITE 390 390 ACYLATED BY PENICILLIN (BY SIMILARITY).  
 SQ SEQUENCE 914 AA; 99562 MW; 6978E33DFE2423E6 CRC64;  
 Query Match 10.1%; Score 86; DB 1; Length 914;  
 Best Local Similarity 23.0%; Pred. No. 49;  
 Matches 45; Conservative 33; Mismatches 56; Indels 62; Gaps 9;  
 QY 3 VKEFLNKDTGSEVSELEKPHRTVTITONGKEMSTIVSEEDFILPYKGELEKGYFDGWE 62  
 DB 737 VKQSV---DGGSYSE-----IONSS-----AEAVISSGVQPSV---YKFEVTA 774  
 QY 63 ISGFEKKDA--GYVINLSKDTFIKPVKKLEKKEENK----- 100  
 DB 775 VSD-DGKSTASTSYEVPRAEDDEDKDDQQTDEKQDDKTQDDTQTDDSGKDDGQTDQD 833  
 QY 101 ----FTFVSKKKDNPQVNHSQLN-----ESIRKEDLQREHSQKSDSTKDYTA 145  
 DB 834 QTDSTNDQDKQDDTNTNPSDNNQDOSNDNDNSNNQDTSDDSGKNDSTG--SD 891  
 QY 146 TVLDKNNSISKSTNN 161  
 DB 892 TNKNTDTSNKQTINS 907  
 RESULT 23  
 MDN1 HUMAN STANDARD; PRT; 5596 AA.  
 ID MDN1 HUMAN  
 AC Q9NUZ2; O15019;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Midasin (MIDAS-containing protein).  
 GN MDN1 OR KIA0301.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX Garbarino J.E., Gibbons I.R.;  
 RT "Expression and genomic analysis of midasin, a novel and highly  
 RT conserved AAA protein distantly related to dynein.";  
 RL BMC Genomics 3:18-18(2002).  
 RN [2]  
 RP SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.  
 RC TISSUE=Testis;  
 RA Tracey A.;  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 3550-5596 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kocani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 4:141-150 (1997).  
 CC -1- FUNCTION: May function as a nuclear chaperone and be involved in  
 CC the assembly/disassembly of macromolecular complexes in the  
 CC nucleus.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: Contains 1 WMPA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



```

Dn 82 SNFGQWNLFLAEGESKNDADAEAFV--KKQVDLDKIIIRKGLVKMAHIDSKQEEFEKE-K-X-133
Qy 105 VSKKDN-----PQVHSQLNESHKREDLQREHSQKSDTKVY 144
Db 139 VEKENDSDDEELMADMGFGWGAFMNNGDENOSHEEHEEHEEKEEHEEHEEBOEEMT 192

RESULT 25
RGAL YEAST STANDARD; PRT; 1007 AA.
AC P39083; P39934;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Rho-type GTPase-activating protein 1.
GN RGAL OR DDM1 OR THE1 OR YOR127W OR C0290 OR YOR3290W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycus.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RC STRAIN=S288C;
RX MEDLINE=96239492; PubMed=8657111;
RA Chen G.-C., Zheng L., Chan C.S.M.;
RT "The LIM domain-containing Dbp1 GTPase-activating protein is required
RL for normal cellular morphogenesis in Saccharomyces cerevisiae.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97060020; PubMed=8904341;
RA Wiemann S., Rechmann S., Benes V., Voss H., Schwager C., Wlecek C.,
RA Stiegemann J., Zimmermann J., Erfle H., Paces V., Ansoerge W.;
RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV
RL from Saccharomyces cerevisiae reveals 30 open reading frames.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97060020; PubMed=8904341;
RA Wiemann S., Rechmann S., Benes V., Voss H., Schwager C., Wlecek C.,
RA Stiegemann J., Zimmermann J., Erfle H., Paces V., Ansoerge W.;
RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV
RL from Saccharomyces cerevisiae reveals 30 open reading frames.";
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=96101594; PubMed=7498791;
RA Stevenon B.U., Ferguson B., de Virgilio C., Bi E., Pringle J.R.,
RA Ammerer G., Sprague G.F. Jr.;
RT "Mutation of RGAL, which encodes a putative GTPase-activating protein
RL for the polarity-establishment protein Cdc42p, activates the
RN pheromone-response pathway in the yeast Saccharomyces cerevisiae.";
RN [5]
RP SEQUENCE OF 570-639 FROM N.A.
RC STRAIN=SNY243;
RX MEDLINE=93087574; PubMed=1454852;
RA Ramer S.W., Elledge S.J., Davis R.W.;
RT "Dominant genetics using a yeast genomic library under the control of
RN a strong inducible promoter.";
CC -1- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR CDC42 AND/OR RHOL.
CC NEGATIVE REGULATOR OF THE PHEROMONE-RESPONSE PATHWAY THROUGH THE
CC SHP2 PROTEIN KINASE; ACTS AT A STEP BETWEEN THE G-PROTEIN AND THE
CC MAP KINASE MODULE. DOMINANT SUPPRESSOR OF BUD EMERGENCE DEFECT
CC CAUSED BY DELETION OF IPL2/BEN2. INVOLVED IN THE CONTROL OF
CC POLARIZED CELL GROWTH AND PROPER BUD SITE SELECTION.
CC -1- SIMILARITY: Contains 2 LIM zinc-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics, and the EMBL outstation

```

[illegible]

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1  
 light chain LC1].  
 GN MAP1B.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RC STRAIN=Sprague-Dawley; TISSUE=Testis;  
 RX MEDLINE=96257242; PubMed=8666295;  
 RA Liu D., Fischer I.;  
 RT "Isolation and sequencing of the 5' end of the rat microtubule-  
 associated protein (MAP1B)-encoding cDNA.";  
 RL Gene 172:307-308 (1996).  
 RN [2]  
 RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Gli1al tumor;  
 RX MEDLINE=92347374; PubMed=1639092;  
 RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;  
 RT "Identification of two distinct microtubule binding domains on  
 recombinant rat MAP 1B.";  
 RL Eur. J. Cell Biol. 57:66-74 (1992).  
 RN [3]  
 RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.  
 RC TISSUE=spinal cord; PubMed=2555150;  
 RX MEDLINE=90059871; PubMed=2555150;  
 RA Rientz A., Greeningloh G., Hermans-Borgmeyer I., Kirsch J.,  
 RA Litzner U.Z., Prior P., Gundelfinger E.D., Schmidt B., Betz H.;  
 RT "Neuraxin, a novel putative structural protein of the rat central  
 nervous system that is immunologically related to microtubule-  
 associated protein 5.";  
 RL EMBO J. 8:2879-2888 (1989).  
 RN [4]  
 RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.  
 RX MEDLINE=97405699; PubMed=9260743;  
 RA Ma D., Nothias F., Boyne L.J., Fischer I.;  
 RT "Differential regulation of microtubule-associated protein 1B (MAP1B)  
 in rat CNS and PNS during development.";  
 RL J. Neurosci. Res. 49:319-332 (1997).  
 CC -1- FUNCTION: The function of brain MAPs is essentially unknown.  
 CC Phosphorylated MAP1B may play a role in the cytoskeletal changes  
 CC that accompany neurite extension. Possibly MAP1B binds to at least  
 CC two tubulin subunits in the polymer, and this bridging of subunits  
 CC might be involved in nucleating microtubule polymerization and in  
 CC stabilizing microtubules.  
 CC -1- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate  
 CC with MAP1A and MAP1B proteins.  
 CC -1- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,  
 CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,  
 CC heart or muscle.  
 CC -1- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic  
 CC nerve levels are high early in development but decrease during  
 CC postnatal development and are low in adults. In dorsal root  
 CC ganglia levels remain high throughout development.  
 CC -1- INDUCTION: By nerve growth factor.  
 CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
 CC responsible for the binding of MAP1B to microtubules.  
 CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated  
 CC from MAP1B by proteolytic processing. It is free to associate with  
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region  
 CC of MAP1B (by similarity).  
 CC -1- PTM: Phosphorylated.  
 CC -1- SIMILARITY: TO MAP1A.  
 CC -1- CAUTION: A C-terminal fragment of this protein (residues 1597 to  
 CC 2459) was originally described as neuraxin in Ref.3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U52950; AAB17068.1; -  
 DR EMBL: X60370; CAC16162.1; -  
 DR EMBL: X16623; CAA34620.1; ALT\_SEQ.  
 DR PIR: A56577; A56577.  
 DR InterPro: IPR00102; MAP1B\_neuraxin.  
 DR Pfam: PRO0414; MAP1B\_neuraxin: 10.  
 DR PROSITE: PS00230; MAP1B\_NEURAXIN; 8.  
 KW Microtubules; Repeat; Phosphorylation.  
 FT CHAIN ? 2459  
 FT REPEAT 1869 1885 MAP1B 1.  
 FT REPEAT 1886 1902 MAP1B 2.  
 FT REPEAT 1903 1919 MAP1B 3.  
 FT REPEAT 1920 1936 MAP1B 4.  
 FT REPEAT 1937 1953 MAP1B 5.  
 FT REPEAT 1954 1970 MAP1B 6.  
 FT REPEAT 1988 2004 MAP1B 7.  
 FT REPEAT 2005 2021 MAP1B 8.  
 FT REPEAT 2022 2038 MAP1B 9.  
 FT REPEAT 2039 2055 MAP1B 10.  
 FT DOMAIN 559 1035  
 FT DOMAIN 588 786  
 FT DOMAIN 2224 2312  
 FT CONFLICT 127 127 M -> V (IN REF. 1).  
 FT CONFLICT 140 140 T -> S (IN REF. 1).  
 FT CONFLICT 2112 2112 R -> K (IN REF. 3).  
 FT CONFLICT 2169 2169 L -> I (IN REF. 3).  
 SQ SEQUENCE 2459 AA; 269497 MW; 2E3F6872DEDB8BA2 CRC64;  
 Query Match 10.1%; Score 85.5; DB 1; Length 2459;  
 Best Local Similarity 24.4%; Pred. No. 1.5e+02;  
 Matches 40; Conservative 26; Mismatches 53; Indels 45; Gaps 8;  
 QY 4 KEFLNKDGEVSELPKRVTVTIQNGKMSSTIVSEEDFLIPYKGELEKGYQFDGWEI 63  
 DB 583 EKVYKADPKGVESKPSSTEKEVSKERQ-----PV-KAIV-----EK 622  
 QY 64 SREGKQAGVYINISKDTFFIKVFK-KLEEKEENKFTFVSKKQDPQVNSQLNES 122  
 DB 623 AATESKP-----KYTKQVKKKIKTKPEKKEE-KRKKEVAKED----- 662  
 QY 123 HKREDLOREHSQKSDSTQVATVLD--DKNNISSKSTNNRNK 164  
 DB 663 --KTPKKDEKPKGEAKKEIKKEIKKEKKELKKEVKKETPLK 704  
 RESULT 27  
 YBLE SCHPO STANDARD; PRT; 719 AA.  
 AC Q10342; Q9UR08;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C106.14c in chromosome II.  
 GN SPBC106.14C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=2148401; PubMed=11859360;  
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayes S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Baker S., Boller T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,



```

RA Chen Z., Wen Y. ;
RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
CC
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.1db-sib.ch/announce/1db-sib.ch).
CC or send an email to license@1db-sib.ch.
CC -----
CC EMBL: AB016747; AAC04542.1; -.
DR HAMAP: MF_00100; -. 1.
DR Pfam: PF00009; GTP_EFTU_1.
DR Pfam: PF03144; GTP_EFTU_D2_1.
DR Pfam: PF04760; IF2_N_2.
DR ProDom: PD186100; IF2_1.
DR TIGRFAMs: TIGR00487; IF-2_1.
DR TIGRFAMs: TIGR00231; eIF2_GTP_1.
DR PROSITE: PS01176; IF2_1.
DR Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 225 373 G-DOMAIN.
FT NP_BIND 231 238 GTP (BY SIMILARITY).
FT NP_BIND 277 281 GTP (BY SIMILARITY).
FT NP_BIND 331 334 GTP (BY SIMILARITY).
SQ SEQUENCE 720 AA; 79343 MW; 07FB5A659CE970C CRC64;

Query Match 10.0%; Score 85; DB 1; Length 720;
Best Local Similarity 28.0%; Pred. No. 45;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2

QY 91 IEEKKEENKPTDPVSKKKD---NPQVNSQLNESHKEDLQREHSQKSDSTK----D 142
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 38 LEEQIQIALDKKFFAASQAKQNTKQNTQNNQKSNKQNSNDKEKQGSKNNSKPTKKKQN 97
QY 143 VTATVLDRNNISSKSTTNPNK 164
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 NKGKQKNKNTKNNQKNNKK 119

RESULT 29
Y040 MYCPN STANDARD; PRT; 657 AA.
AC P75062;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipidprotein MG040 homolog precursor (D09_orf657).
GN MPM052 OR MPl02.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
[1]
RA SEQUENCE FROM N.A.
RA STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; Pubmed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pilkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
CC Nucleic Acids Res. 24:4420-4449 (1996).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (potential).
CC -!- SIMILARITY: SOME, TO T.PALLIDIUM TMAPC.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: A5000012; AAB95750.1; -.  
DR PIR: S73428; S73428.  
DR InterPro: IPR003760; Bmp.  
DR Pfam: PF02608; Bmp; 1.  
DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Hypothetical protein; Lipoprotein; Membrane; Signal;  
KW Complete proteome.  
FT SIGNAL 1 26 POTENTIAL.  
FT CHAIN 27 657 HYPOTHETICAL\_LIPOPROTEIN MG040 HOMOLOG.  
FT LIPID 27 27 N-ACYL DIGLYCERIDE (POTENTIAL).  
SQ SEQUENCE 657 AA; 71671 MW; 6147815673606A9 CRC64;  
  
Query Match 10.0%; Score 84.5; DB 1; Length 657;  
Best Local Similarity 16.7%; Pred. No. 44;  
Matches 42; Conservative 36; Mismatches 78; Indels 95; Gaps 5;  
  
QY 8 LNKDTEGSELKPRVTVTIONGKEM-----SSTVSEE 41  
DB 343 VLLANVEVARRKHTALIGVDSAGELLDINQAPDQILKSKKIIPSSIKALDAIE 402  
QY 42 DFLPYKGELEKGYOPDG-----WEISGEFGKDAY 74  
DB 403 NMLIALIQKSDNNNGYGFYNNIGVTSSVGISEAGYEPLDIPVWKTTQSGKSMATN 462  
QY 75 VINL-----SKPTIKVFPEKIEKE----- 96  
DB 463 MNLKLLSSDDTNTKALKREVSTHKGSGDKGIIGKYSNLLTKSTTTVAQKSWTNN 522  
QY 97 ---EENKPTFDV-SKKDNPQVNSQLNESHREKEDQREHSQKSDSTQVATVLDKNN 152  
DB 523 SGTEQKKNSEVDTKKKEKSKKTOSNGDSQQNGKEINDIISKSLTITMTMANNV 582  
QY 153 ISSKSTNNPN 163  
DB 583 MSSKQSSDN 593  
  
RESULT 30  
ID IGA0\_HAEIN STANDARD; PRT; 1694 AA.  
AC P44969;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAI protease).  
GN IGA OR IGA1 OR HI0990.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxId=727;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=serotype D;  
RA Wright A., Fishman Y., Tai F., Plaut A.G.;  
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.  
RN 12  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KM20 / ATCC 51907;  
RX MEDLINE=95350630; Pubmed=7542800;  
RA Fleischmann A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick E.F.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.M.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedlorn E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd".  
RL Science 269:496-512(1995).  
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
CC -1- PRODUCING INTRACT FC AND FAB FRAGMENTS.  
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
CC certain Pro-Xaa bonds in the hinge region. No small molecule  
CC substrates are known.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X59800; -; NOT ANNOTATED\_CDS.  
DR EMBL: U32779; AAC22651.1; -.  
DR PIR: H64106; H64106.  
DR MEROPS: S06.001; -.  
DR TIGR: H10990; -.  
DR InterPro: IPR006315; Autotransport.  
DR InterPro: IPR005546; Autotransporter.  
DR InterPro: IPR000710; IGA\_S6.  
DR InterPro: IPR004899; Pertactin.  
DR Pfam: PF03797; Autotransporter; 1.  
DR Pfam: PF02395; IGA1; 1.  
DR Pfam: PF0312; Pertactin; 1.  
DR PRINTS: PR00921; IGASERPTASE.  
DR TIGRPMs: TIGR01414; autotrans\_bar1; 1.  
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal;  
KW Complete proteome.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.  
FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).  
FT ACT\_SITE 288 288 PROBABLE.  
FT CONFLICT 253 254 EN -> GV (IN REF. 1).  
FT CONFLICT 272 272 G -> A (IN REF. 1).  
FT CONFLICT 464 464 S -> E (IN REF. 1).  
FT CONFLICT 866 866 S -> T (IN REF. 1).  
FT CONFLICT 1036 1036 A -> D (IN REF. 1).  
FT CONFLICT 1074 1074 A -> G (IN REF. 1).  
FT CONFLICT 1421 1421 A -> G (IN REF. 1).  
FT CONFLICT 1545 1545 H -> T (IN REF. 1).  
SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;  
  
Query Match 10.0%; Score 84.5; DB 1; Length 1694;  
Best Local Similarity 26.1%; Pred. No. 1.2e+02;  
Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;  
  
QY 76 INLSKDTFKIPFKLIEKEEENKPTFVSKKDNPQVNSQLNESHREKEDQREHS- 134  
DB 1288 INTGSATALTETAEBSKDKQTETASTEDASQKAKYTVADNSVANNSSSDPKSRRRI 1347  
QY 135 --QKSDSTQVATVLDKNNISSKSTNNPNK 164  
DB 1348 SQPETSARETTASTDETTIADNSKSKMNR 1379  
  
RESULT 31  
ID HG2\_HALRO STANDARD; PRT; 510 AA.  
HG2\_HALRO

AC Q02508; 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Protein HG2.  
GN HG2.  
OS Halocynthia roretzi (Sea squirt).  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
OC Stolidobranchia; Pyrosidae; Halocynthia.  
OX NCBI\_TaxID=7729;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Oocyte;  
RA MEDLINE=93224498; PubMed=8468323;  
RA Fujiwara S., Kawahara H., Makabe K.W., Satoh N.;  
RT "A complementary DNA for an ascidian embryonic nuclear antigen Hg2  
RT encodes a protein closely related to the amphibian histone-binding  
RT protein N1.";  
RL J. Biochem. 113:189-195(1993).  
CC -1- FUNCTION: MAY FUNCTION AS A NUCLEOSOME ASSEMBLY FACTOR DURING  
CC RAPID EMBRYONIC CELL DIVISIONS.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: EMBRYO AND LARVAE.  
CC -1- DEVELOPMENTAL STAGE: THIS PROTEIN IS DETECTED IN THE NUCLEI  
CC OF ALL CELLS IN EMBRYOS AND LARVAE BUT IS NOT DETECTED IN THE  
CC CELLS OF METAMORPHOSED JUVENILES.  
CC -1- SIMILARITY: TO XENOPUS LAEVIS HISTONE BINDING PROTEIN N1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D13541; BAA02741.1; -  
DR PIR; JX0254; JX0254.  
DR InterPro; IPR001440; TPR.  
DR Pfam; PF00515; TPR; 2.  
KW Nuclear protein; Developmental protein.  
FT DOMAIN 98 143 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 219 238 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 444 451 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 465 471 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
SQ SEQUENCE 510 AA; 56871 MW; D9961E5953E976FF CRC64;  
Query Match 9.9%; Score 84; DB 1; Length 510;  
Best Local Similarity 28.4%; Pred. No. 36;  
Matches 50; Conservative 25; Mismatches 61; Indels 40; Gaps 12;  
QY 4 KEFILN-KDTGSEVSELKPHRTVTITQNGKMSSTVSEEDFLPYKGELEKGYFDGWE 62  
DB 354 KEIISEKVEGELKEIP-----DINSKI--BDVILA--KKQMK--LDG-- 392  
QY 63 ISGF---EGKDAQVIVL--SKDTPIK--PVK---KIEEKKEENKPTPVYSKKKNP 112  
DB 393 -SPFQASGESSSGGASTSDKPCSTPIPIKVAPTSPVAKDPSDITLHVRKRPS 451  
QY 113 ---OVNHSQLESRHKEDLQREHSQKSDSTKDTATATLDKNNISKSTNNPNK 164  
DB 452 DEDNQPAESKESKAKAKQEBTEATNGHSAVKQOT-DVTDKNGNGHSGKT--PKK 504  
RESULT 32  
RBP2\_PLAVB STANDARD; PRT; 1251 AA.  
AC Q00759;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Reticulocyte binding protein 2 (Fragment).  
GN RBP2.

OS Plasmodium vivax (strain Belen).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=31273;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=92315338; PubMed=1617731;  
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
RT "A reticulocyte-binding protein complex of Plasmodium vivax  
RT merozoites.";  
RL Cell 69:1213-1226(1992).  
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
CC HUMAN RETICULOCYTE CELLS.  
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M88098; AAA29744.1; -  
KW Malaria; Receptor; Membrane.  
FT NON\_TER 1 1  
FT NON\_TER 1 1  
SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;  
Query Match 9.9%; Score 84; DB 1; Length 1251;  
Best Local Similarity 23.0%; Pred. No. 96;  
Matches 42; Conservative 37; Mismatches 72; Indels 32; Gaps 9;  
QY 4 KEFILNKD--TGEVSELKPH--RYVTITQNGKMSSTI-----VSEEDFLPYKGE 52  
DB 253 REIMMKKQIKSYSLSEIKYKDKCTEISNKGKDKLEFKRPNSESNKNKVINET 312  
QY 53 EKGYPDGMSEISGF--GKDAQVIVL--SKDTPIKPVFK---IEEKKEENKPTFD 104  
DB 313 NENINNSQYKLDIEDAKQKSTKVELFKHKEITTSNIFKSEILGVETSKQKINKAD 372  
QY 105 VSK--KKDNPOV-----NHSQLESRHKEDLQREHSQKSDSTKDTATATLD--KN 151  
DB 373 IMKEIERNSSEIQGVKFGQENLKNLNEPHYDNADELDNDKSTNAKVLTETNLESVGH 432  
QY 152 NIS 154  
DB 433 NLS 435  
RESULT 33  
FENR\_PEA STANDARD; PRT; 360 AA.  
AC P10933;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Ferredoxin-NADP reductase, leaf isozyme, chloroplast precursor  
DE (EC 1.18.1.2) (FNR).  
GN PETH.  
OS Pisum sativum (Garden pea).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eufrosids 1; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
OX NCBI\_TaxID=3888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Little Marvel; TISSUE=leaf;  
RA Newman B.O., Gray J.C.;  
RT "Characterisation of a full-length cDNA clone for pea ferredoxin-NADP+  
RT reductase.";  
RL Plant Mol. Biol. 10:511-520(1988).  
RN [2]  
SQ SEQUENCE OF 270-360 FROM N.A., AND MUTAGENESIS.

RX MEDLINE=93374905; PubMed=8366077;  
 RA Orellano E.G., Calcaterra N.B., Carrillo N., Ceccarelli E.A.;  
 RT "Probing the role of the carboxyl-terminal region of ferredoxin-NADP+  
 reductase by site-directed mutagenesis and deletion analysis.";  
 RL J. Biol. Chem. 268:19267-19273(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.70 ANGSTROMS).  
 RA MEDLINE=99396739; PubMed=10467097;  
 RA Deng Z., Alliverti A., Zanetti G., Arakaki A.K., Octado J.,  
 RA Orellano E.G., Calcaterra N.B., Ceccarelli E.A., Carrillo N.,  
 RA Karplus P.A.;  
 RT "A productive NADP+ binding mode of ferredoxin-NADP+ reductase  
 revealed by protein engineering and crystallographic studies.";  
 RL Nat. Struct. Biol. 6:847-853(1999).  
 CC -1- FUNCTION: MAY PLAY A KEY ROLE IN REGULATING THE RELATIVE AMOUNTS  
 CC OF CYCLIC AND NON-CYCLIC ELECTRON FLOW TO MEET THE DEMANDS OF THE  
 CC PLANT FOR ATP AND REDUCING POWER.  
 CC -1- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized  
 CC ferredoxin + NADPH.  
 CC -1- COFACTOR: FAD.  
 CC -1- PATHWAY: FINAL STEP IN LINEAR PHOTOSYNTHETIC ELECTRON TRANSPORT  
 CC CHAIN; IT HAS ALSO BEEN IMPLICATED IN CYCLIC ELECTRON FLOW AROUND  
 CC PHOTOSYSTEM I, ITS ROLE BEING TO RETURN ELECTRONS FROM FERREDOXIN  
 CC TO THE CYTOCHROME B-F COMPLEX.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST; STROMAL SIDE OF THE THYLAKOID  
 CC MEMBRANE IN THE VICINITY OF THE PHOTOSYSTEM I IN THE NON-STACKED  
 CC AND FRINGE PORTION OF THE MEMBRANE.  
 CC -1- MISCELLANEOUS: FNR IS PROBABLY ATTACHED TO THE MEMBRANE BY A  
 CC SPECIFIC BINDING PROTEIN.  
 CC -1- SIMILARITY: WITH OTHER SPECIES FNR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; X12446; CAA30978.1; -;  
 CC DR EMBL; L15565; AAB59349.1; -;  
 CC DR EMBL; L15567; AAB59303.1; -;  
 CC DR EMBL; L15569; AAB59304.1; -;  
 CC DR PIR; S04030; S04030.  
 CC DR PDB; 1OF2; 01-SEP-99.  
 CC DR PDB; 1OFY; 01-SEP-99.  
 CC DR PDB; 1OG0; 01-SEP-99.  
 CC DR PDB; 1QGA; 01-SEP-99.  
 CC DR InterPro; IPR001709; FPN\_cyt\_reductse.  
 CC DR InterPro; IPR001433; Oxred\_FAD/NAD(p).  
 CC DR Pfam; PF00175; NAD\_binding\_1; 1.  
 CC DR PRINTS; PR00371; FPNCR.  
 CC DR Oxidoreductase; Flavoprotein; NADP; FAD; Chloroplast; Transit peptide;  
 CC Electron transport; Photosynthesis; Thylakoid; Membrane;  
 CC Multigene family; 3D-structure.  
 CC KM TRANSIT 1 52 CHLOROPLAST.  
 CC FT 53 360 FERREDOXIN-NADP REDUCTASE, LEAF ISOZYME.  
 CC FT 178 183 BY SIMILARITY.  
 CC FT NP\_BIND 212 230 NADP (RIBOSE PART) (BY SIMILARITY).  
 CC FT MUTAGEN 360 360 Y-W,F,S.  
 CC FT STRAND 68 68  
 CC FT STRAND 71 72  
 CC FT STRAND 78 78  
 CC FT STRAND 80 81  
 CC FT STRAND 84 93  
 CC FT TURN 97 98  
 CC FT STRAND 103 109  
 CC FT TURN 111 112  
 CC FT STRAND 118 119  
 CC FT STRAND 121 125  
 CC FT STRAND 129 129  
 CC FT TURN 131 132

FT STRAND 135 135  
 FT STRAND 139 143  
 FT TURN 147 148  
 FT TURN 150 151  
 FT STRAND 156 162  
 FT STRAND 165 167  
 FT TURN 169 170  
 FT STRAND 173 175  
 FT STRAND 177 184  
 FT TURN 187 188  
 FT STRAND 190 197  
 FT TURN 200 201  
 FT STRAND 204 204  
 FT TURN 207 208  
 FT STRAND 210 216  
 FT STRAND 217 220  
 FT HELIX 221 231  
 FT HELIX 232 232  
 FT STRAND 243 250  
 FT STRAND 253 255  
 FT TURN 257 258  
 FT HELIX 259 268  
 FT TURN 270 272  
 FT STRAND 273 279  
 FT STRAND 280 282  
 FT STRAND 284 284  
 FT TURN 287 287  
 FT STRAND 290 290  
 FT STRAND 293 297  
 FT HELIX 298 300  
 FT HELIX 301 309  
 FT TURN 311 312  
 FT STRAND 313 319  
 FT TURN 321 322  
 FT HELIX 323 337  
 FT TURN 338 339  
 FT HELIX 342 351  
 FT TURN 352 353  
 FT STRAND 355 360  
 SQ SEQUENCE 360 AA; 40194 MW; 7F1C10DEBBA7B24 CRC64;  
 Query Match 9.8%; Score 83.5; DB 1; Length 360;  
 Best local Similarity 24.7%; Pred. No. 27;  
 Matches 38; Conservative 19; Mismatches 50; Indels 47; Gaps 7;  
 QY 3 VKEPIINKDTGEV-----SELKP-HRVVTIIONGKEM-----SSTI 37  
 Db 161 VKRLVYTNDAAGVVGVCNSFLCDLPGSEVKITGPVGKEMMPKDPNATVIMLTGTGTGI 220  
 QY 38 VEEEDPILPVYKGELEKGYQPDG--WEISGFEGKKDAGVIVNLSKDTFIKPVFKKIEEK 95  
 Db 221 APPRSFLTMWF-PEKHEDYQFNGMLFLG-----VPTSSSLYKEFEKKEKKA 269  
 QY 96 EENKPTPDVSKKKONPVNHSQLNESHREKDLQ 129  
 Db 270 PENFRIDFAVRSR-----QVNDKGEKWIQ 294  
 RESULT 34  
 ID \_U2RL HUMAN STANDARD; PRT; 479 AA.  
 AC Q15655; Q13570;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit  
 DE related-protein 1.  
 GN UZAF1-RS1 OR UZAF1RS1 OR UZAFBPL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN NCBI\_TaxId=9606; [1]



[3]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-Placenta;  
 RX MEDLINE=89343979; PubMed=2527334;  
 RA Hickey E., Brandon S.B., Smale G., Lloyd D., Weber L.A.;  
 RT "Sequence and regulation of a gene encoding a human 89-kilodalton  
 RT heat shock protein.";  
 RT Mol. Cell. Biol. 9:2615-2626(1989).  
 [4]  
 RN SEQUENCE OF 1-311 FROM N.A.  
 RP MEDLINE=88056312; PubMed=2445630;  
 RA Hoffmann T., Hovemann B.;  
 RT "Cloning and nucleotide sequence of the murine hsp84 cDNA and  
 RT chromosome assignment of related sequences.";  
 RT Gene 56:29-40(1987).  
 [5]  
 RN SEQUENCE OF 1-311 FROM N.A.  
 RP MEDLINE=90076956; PubMed=2591742;  
 RA Walter T., Drabant B., Krebs H., Tomalak M., Hees S.,  
 RA Benecke B.J.;  
 RT "Cloning and analysis of a human 86-kDa heat-shock-protein-encoding  
 RT gene.";  
 RT Gene 83:105-115(1989).  
 [6]  
 RN SEQUENCE OF 184-731 FROM N.A.  
 RP TISSUE-Placenta;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Mansura K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [7]  
 RN SEQUENCE OF 538-731 FROM N.A.  
 RP TISSUE-Heart;  
 RA Tanaka M., Tanaka T., Mitsu Y., Yamamoto M., Wood J.N.;  
 RL Submitted (SBP-1996) to the EMBL/GenBank/DBJ databases.  
 [8]  
 RN SEQUENCE OF 1-20, AND PHOSPHORYLATION.  
 RX MEDLINE=89123325; PubMed=2492519;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "Two human 90-kDa heat shock proteins are phosphorylated in vivo at  
 RT conserved serines that are phosphorylated in vitro by casein kinase  
 RT II.";  
 RT J. Biol. Chem. 264:2431-2437(1989).  
 [9]  
 RN PHOSPHORYLATION BY DS-DNA KINASE.  
 RX MEDLINE=90008887; PubMed=2507541;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "The human double-stranded DNA-activated protein kinase phosphorylates  
 RT the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal  
 RT threonine residues.";  
 RT J. Biol. Chem. 264:17275-17280(1989).  
 [10]  
 RN INTERACTION WITH OM34.  
 RP MEDLINE=98324997; PubMed=9660753;  
 RA Young J.C., Obermann W.M., Hartl F.U.;  
 RT "Specific binding of tetratricopeptide repeat proteins to the

RT C-terminal 12-kDa domain of hsp90.";  
 RL J. Biol. Chem. 273:18007-18010(1998).  
 [11]  
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 10-222.  
 RP MEDLINE=97262065; PubMed=9108479;  
 RA Stebbins C.E., Russo A.A., Schneider C., Rosen N., Hartl F.U.,  
 RA Pavletich N.P.;  
 RT "Crystal structure of an Hsp90-geldanamycin complex: targeting of a  
 RT protein chaperone by an antitumor agent.";  
 RT Cell 89:239-250(1997).  
 [12]  
 RN X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 10-222.  
 RP MEDLINE=99034582; PubMed=9817749;  
 RA Obermann W.M., Sondermann H., Russo A.A., Pavletich N.P., Hartl F.U.;  
 RT "In vivo function of Hsp90 is dependent on ATP binding and ATP  
 RT hydrolysis.";  
 RT Cell Biol. 143:901-910(1998).  
 CC -1- FUNCTION: Molecular chaperone. Has ATPase activity  
 CC (By similarity).  
 CC -1- SUBUNIT: Homodimer. Interacts with OM34.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X15183; CAA33259.1; -  
 DR EMBL: X07270; CAA30255.1; -  
 DR EMBL: M27024; AAA63194.1; -  
 DR EMBL: M30626; AAA56023.1; -  
 DR EMBL: BC000987; AAH00987.1; -  
 DR EMBL: D87666; BAA13430.1; -  
 DR EMBL: D87666; BAA13431.1; -  
 DR PIR: A32319; HHHU86.  
 DR PDB: 1YER; 22-APR-98.  
 DR PDB: 1YES; 22-APR-98.  
 DR PDB: 1BYO; 28-OCT-98.  
 DR TRASNFA: T00992; -  
 DR Gene: HGNC:5253; HSPCA.  
 DR MIM: 140571; -  
 DR GO: GO:0005737; C:cytoplasm; NAS.  
 DR GO: GO:0003773; F:heat shock protein activity; NAS.  
 DR InterPro: IPR003594; AtPbind\_Arpase.  
 DR InterPro: IPR001404; Hsp90.  
 DR Pfam: PF02518; HATPase\_C; 1.  
 DR Pfam: PF00183; HSP90; 1.  
 DR PRINTS: PR00775; HEATSHOCK90.  
 DR SMART: SM00387; HATPase\_C; 1.  
 DR PROSITE: PS00298; HSP90; 1.  
 KW Chaperone; ATP-binding; Heat shock; Phosphorylation; 3D-structure.  
 FT INIT MET 0 0  
 FT MOD RES 4 4  
 FT MOD RES 6 6  
 FT MOD RES 230 230  
 FT MOD RES 262 262  
 FT CONFLICT 62 62  
 FT STRAND 17 20  
 FT HELIX 23 34  
 FT TURN 40 41  
 FT HELIX 42 64  
 FT HELIX 66 69  
 FT TURN 70 71  
 FT STRAND 77 82  
 FT TURN 83 86  
 FT TURN 87 92  
 FT HELIX 99 103  
 FT TURN 104 107  
 T -> S (IN REF. 3, 4 AND 5).

```

FT HELIX 110 122
FT TURN 123 124
FT HELIX 127 133
FT TURN 134 134
FT HELIX 136 142
FT TURN 143 143
FT STRAND 144 152
FT TURN 154 155
FT STRAND 158 163
FT TURN 166 167
FT STRAND 168 173
FT STRAND 182 189
FT HELIX 191 197
FT HELIX 199 209
FT TURN 211 212
FT STRAND 217 219
SQ SEQUENCE 731 AA; 84542 MW; 1249ABCFCE06297C CRC64;

Query Match 9.8%; Score 83.5; DB 1; Length 731;
Best Local Similarity 23.6%; Pred. No. 58;
Matches 39; Conservative 28; Mismatches 59; Indels 39; Gaps 7;

OY 6 FILAKDTGSEVSEILKPRVVTITONCKEMSSIVSEDFILPYKGELEKGYQFDCWEISG 65
DB 169 FTVRTDTGE---PMGRGTKVILHLKEDQTEYLEERRI-----KEIVKHSQFIDGPIITL 219
OY 66 F-EGKKDAGYVNIISKDTFIKVPFKKIEKKEENKPTFDVSKKKONPOVNHSHR 124
DB 220 FVEKERDK---EVSDD-----EAEKEDEKEKEKEKESESDKPEI----- 257
OY 125 KEDLQREHSHQKSDSTKDTATVLDKN-----NISSKSTTNPN 163
DB 258 -EDVGSDEEEKKDDKKKKKKIKERYIDQELNKTPIWTNP 301

RESULT 37
ARS2_DROME
ID ARS2_DROME STANDARD; PRT; 943 AA.
AC OY9VK7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arsenite-resistance protein 2 homolog.
GN CG7843.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Abmayant A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brockstein P., Broctier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Caceres A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster K., Gabbitarian A.E., Gary N.S., Gelbart W.M., Glaser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houston D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

```

```

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weissstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RC STRAIN=Berkley;
RX MEDLINE=22426069; PubMed=12537572;
RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Bernan B.P.,
RA Betrencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter U., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q9V9K7-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q9V9K7-2; Sequence=VSP 000327;
CC -I- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, AE003784; AAM68343.1; -
CC DR EMBL, AE003784; AAM68345.1; -
CC DR Flybase; FBgn003062; CG7843.
CC DR Pfam; PF04959; ARS2; 1.
CC Hypothetical protein; Alternative splicing.
CC VARSPLIC 47 50 Missing (in isoform short).
CC /FtId=VSP 000327
SQ SEQUENCE 943 AA; 107221 MW; 0C1AF09E02BEA80B CRC64;

Query Match 9.8%; Score 83.5; DB 1; Length 943;
Best Local Similarity 25.0%; Pred. No. 77;
Matches 36; Conservative 23; Mismatches 68; Indels 17; Gaps 4;

OY 14 EYSELKP-----HRYVTITONCKEMSSIT-----VEEDFILPYKGELEKGYQFDCW 61
DB 272 KYLEKPDPPVYEKAKOMQSVKEVETIKSPKEKSEADPVSTQRPVPPVNSDGENW 331
OY 62 EISGEGKADAGYVNIISKDTFIKVPFKKIEKKEENKPTFDVSKKKONPOVNHSHR 118
DB 332 DDDDAENAPPKKELAEKSDSKSDPKQJLKKTKKKRSSDDSSSSSSSSSSSSDEEK 391
OY 119 LNESHKEDLQREHSHQKSDSTKD 142
DB 392 LKEKYVDVGDGRAE--QKTEAKND 413

```





MEMBRANE PROTEINS. REQUIRED FOR HEME UTILIZATION AND VIRULENCE.  
 -1- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE PERIPLASM (BY SIMILARITY).  
 -1- SIMILARITY: BELONGS TO THE TONB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U04996; AAA60460.1; -  
 CC EMBL; U32711; AAC21917.1; -  
 CC PIR; F64057; F64057.  
 CC TIGR; H10251; -  
 CC InterPro; IPR003538; TonB\_C.  
 CC InterPro; IPR006250; TonB\_C.  
 CC Pfam; PF03544; TonB; 1.  
 CC PRINTS; PR01374; TONBPROTEIN.  
 CC TIGRPFAM; TIGR01352; tonb Cterm, 1.  
 CC Transport; Protein transport; Inner membrane; Periplasmic; Transmembrane; Signal-anchor; Repeat; Virulence; Complete proteome.  
 FT DOMAIN 1 7  
 FT TRANSMEM 8 27  
 FT DOMAIN 28 270  
 FT DOMAIN 70 77  
 FT DOMAIN 58 90  
 FT DOMAIN 98 143  
 FT VARIANT 17 17  
 FT VARIANT 31 31  
 FT VARIANT 66 66  
 FT VARIANT 76 84  
 FT VARIANT 116 116  
 FT VARIANT 120 120  
 FT VARIANT 134 134  
 FT VARIANT 140 140  
 FT VARIANT 146 146  
 FT VARIANT 226 226  
 FT VARIANT 226 226  
 FT SEQUENCE 270 AA; 29193 MW; 4CD54B46F7D10A0 CRC64;  
 Query Match 9.8%; Score 83; DB 1; Length 270;  
 Best Local Similarity 32.1%; Pred. No. 22;  
 Matches 26; Conservative 16; Mismatches 33; Indels 6; Gaps 3;  
 QY 80 KDTFKPVFKTIEKKKEENKPTFDVSKKKDNPVNHSQLNESHKREDLQREHGXKSDS 139  
 DB 93 EDPFKPEPKKKEPEKPEKPKPK--GKPKGKPK-NKPKKEVYKPKKPKPKKE--LPKGD 146  
 QY 140 TKDVATVLDKNNISSKSTTN 160  
 DB 147 NIDSSANVNDKASTTSANSN 167  
 RESULT 40  
 ADVS\_HUMAN STANDARD; PRT; 715 AA.  
 ID ADVS\_HUMAN  
 AC 09Y6U3; O8WU97;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-EB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Adseverin (Scinderin).  
 GN SCIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=skin;  
 RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins R.S., Wagner L., Shemmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheef C.F., Blat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Malahy S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE OF 1-527 FROM N.A.  
 RA Kalicki J., Smith-Craig R.,  
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 248-715 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nakakura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.,  
 RT "NEO human cDNA sequencing project."  
 RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: Ca(2+)-dependent actin filament severing protein that is  
 CC presumed to have a regulatory function in exocytosis by affecting  
 CC the organization of the microfilament network underneath the  
 CC plasma membrane. In vitro, also has barbed end capping and  
 CC nucleating activities in the presence of Ca(2+).  
 CC -1- SIMILARITY: BELONGS TO THE VILLIN/GELSOLIN FAMILY.  
 CC -1- SIMILARITY: Contains 6 gelsolin-like repeats.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; BC021090; AAH21090.1; -  
 CC EMBL; AC005281; AAD15423.1; -  
 CC EMBL; AK027778; BAB5361.1; -  
 CC HSSP; P02640; 2VIL.  
 DR InterPro; IPR001974; Gelsolin.  
 DR Pfam; PF00626; Gelsolin; 6.  
 DR PRINTS; PR00597; GELSOLIN.  
 DR SMART; SM00262; GEL; 6.  
 KW Cytoskeleton; Actin-binding; Repeat; Calcium; Capping protein.  
 DR ACTIN-SEVERING (POTENTIAL).  
 FT DOMAIN 1 363  
 FT REPEAT 364 715  
 FT REPEAT 27 76  
 FT REPEAT 148 188  
 FT REPEAT 265 307  
 FT REPEAT 398 451  
 FT REPEAT 523 564  
 FT REPEAT 626 668  
 FT SITE 112 119  
 FT SITE 138 146  
 FT SITE 61 61  
 FT CONFLICT 61 61 R -> H (IN REF. 2).  
 FT SEQUENCE 715 AA; 80508 MW; 45FBA2C8CBFDD80 CRC64;

Query Match 9.8%; Score 83; DB 1; Length 715;  
Best Local Similarity 23.1%; Pred. No. 62;  
Matches 34; Conservative 30; Mismatches 47; Indels 36; Gaps 5;

Qy	22	RYVTYTIQNGKEMSSITVSEEDFILPYKGELEKGYQFDGWEISGFEGK-----KDAGYV	75
Db	262	RYTVVAEENPFMSMAMLLSECFILD--HGAAKQIFVWKGKMANPQERKAAMKTAEEFLQO	319
Qy	76	INLSKDTFIK-----PVEKK-----IREKKEENKPTFDVSKKK	109
Db	320	MNYSKNTQIQVLEPGGETPIFKQFQKMDKQSDSGFKVYTEKVAQIKQIIPDASKLH	379
Qy	110	DNPQV--NHSOLNESHKEDLOREHS	134
Db	380	SSPQMAAQHMMVDDSGSKVEIMRVENN	406

Search completed: February 10, 2004, 10:49:55  
Job time : 9.62413 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 8.09802 Seconds  
(without alignments)  
1011.574 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897

Sequence: 1 KIVKDFPARNITVKEFLNK.....ATVLDKNNISSKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100	11.1	1875	1	MLP1_YEAST
2	99	11.0	1398	1	Q02455 saccharomyc
3	97.5	10.9	1345	1	10P2_PLAIFK
4	97.5	10.9	1345	1	YH00_YEAST
5	96	10.7	893	1	MAPB_HUMAN
6	95	10.6	348	1	GYRA_STAEF
7	95	10.6	1202	1	CYL2_HUMAN
8	93.5	10.4	649	1	DPOW_ASCIM
9	93	10.4	715	1	HS70_PYRSA
10	93	10.4	2464	1	ADSV_HUMAN
11	92.5	10.3	443	1	MAPB_MOUSE
12	92.5	10.3	1702	1	DACA_BACSU
13	92.5	10.3	1888	1	IGA2_HAEIN
14	92.5	10.3	406	1	YDR2_SCHPO
15	92	10.3	1332	1	NSH1_MOUSE
16	90	10.0	688	1	SPT7_YEAST
17	90	10.0	778	1	LIP_STAEF
18	89.5	10.0	433	1	YFR8_YEAST
19	89.5	10.0	1066	1	TIG_STEAM
20	89.5	10.0	1694	1	PIK1_YEAST
21	88	9.8	853	1	YCG1_YEAST
22	87.5	9.8	360	1	YCG1_YEAST
23	87.5	9.8	655	1	FENR_PEA
24	87.5	9.8	655	1	SKM1_YEAST
25	87.5	9.8	879	1	GARP_PLAIF
26	87	9.7	6632	1	RA50_PYRHO
27	86.5	9.6	1006	1	UN89_CAEEL
28	86.5	9.6	1251	1	RAT1_YEAST
29	86	9.6	200	1	RBP2_PLAIV
30	86	9.6	443	1	SYST_LYCES
31	86	9.6	914	1	GLNA_PYRKO
32	86	9.6	5596	1	PBPA_BACSU
33	85.5	9.5	752	1	MDN1_HUMAN
					DRS1_YEAST

34	85.5	9.5	1007	1	RGAL_YEAST
35	85.5	9.5	2459	1	MAPB_RAT
36	85	9.5	365	1	FENR_MESCR
37	85	9.5	472	1	6PBD_LACIA
38	85	9.5	657	1	Y040_MYCPN
39	85	9.5	719	1	YBLE_SCHPO
40	85	9.5	720	1	IF2_STAEF
41	85	9.5	895	1	RA50_THEVO
42	84.5	9.4	950	1	IF2_LACIC
43	84	9.4	510	1	HGV2_HALFO
44	84	9.4	796	1	YFC3_YEAST
45	83.5	9.3	363	1	FENR_VICPA

## ALIGNMENTS

RESULT 1  
ID MLP1\_YEAST STANDARD; PRT; 1875 AA.  
AC Q02455;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin-like protein MLP1.  
GN MLP1 OR YKR095W OR YKR415.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c;  
RA MEDLINE=93247549; PubMed=8483450;  
RA Koeilling R., Nguyen T., Chen E.Y., Botstein D.;  
RT "A new yeast gene with a myosin-like heptad repeat structure."; Mol. Genet. 237:359-369(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=94205265; PubMed=8154186;  
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,  
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;  
RT "The complete sequence of a 15,820 bp segment of Saccharomyces cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three new open reading frames.";  
RL Year 9:1349-1354(1993).  
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA REPAIR.  
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.  
CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MLP1".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, L01992; AAA34783.1; -;  
DR EMBL, X73541; CAAS1948.1; -;  
DR EMBL, Z28320; CAAB2174.1; -;  
DR PIR, S38173; S38173.  
DR SGD, S0001803; MLP1.  
DR GO, GO:0005635; C:nuclear membrane; IDA.  
DR GO, GO:0005654; C:nucleoplasm; IDA.  
DR GO, GO:0006606; P:protein-nucleus import; IDA.  
KW Coiled coil; DNA repair.  
FT DOMAIN 69 487 COILED COIL (POTENTIAL).  
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).  
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).  
FT CONFLICT 301 301 R -> A (IN REF. 1).  
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C906667 CRC64;

```

Query March 100; DB 1; Score 100; 11.1%; Score 100; DB 1; Length 1875;
Best Local Similarity 24.4%; Pred. No. 14;
Matches 50; Conservative 38; Mismatches 71; Indels 46; Gaps 10;

Qy 4 VKDFARNITVAKFFILNKDGTGEVSELKPRVTVITLQNGKMSSTVISEEDF-----ILPVY 58
Db ISQTRRETEENWILLNK---EIQDLYDSKSDISIKLGKESRSRIALAEFKILSNITLDT 703
Qy 59 KGE---LEKGVQFDMWEISGEGKKDA-----GYVNLKSDTIFKRVFKKIEK----- 104
Db 704 KAENDQLEKRPDY-----LQNTILKQDSKTHETLNEYVSCSKLSTIVETLLNKEEQKLR 759
Qy 105 -----KEENKPFPTDVSKKKDNQVHNSQNLNESHK-KEDL---QREHSQKSDSTQDV 153
Db 760 VHEKNGLKQELNK-----LSPKEDSLRIWVTOQLQTKEREDLLEBTRKSCQKIDELEDA 815
Qy 154 TA-----TVLDKNNISSKSTNNPN 173
Db 816 LSELKFKTSQKDHHTKQLEEDNNSN 840

RESULT 2
TOP2_PLAFK STANDARD; PRT; 1398 AA.
ID TOP2_PLAFK
AC P41001;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
NCBI_TaxID=5839;
[1]
SEQUENCE FROM N.A.
MEDLINE=94316496; PubMed=8041616;
Cheesman S., McAliese S., Goman M., Johnson D., Horrocks P.,
Ridley R.G., Kilbey B.J.;
"The gene encoding topoisomerase II from Plasmodium falciparum.";
Nucleic Acids Res. 22:2547-2551(1994).
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
BRASSAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
MAKES DOUBLE-STRAND BREAKS.
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
of double-stranded DNA.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
RELAX ONLY NEGATIVE SUPERCOILS.
CC -1- SIMILARITY: Belongs to the type II topoisomerase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; X79345; -; NOT_ANNOTATED_CDS.
DR HSSP; P06786; IBMG.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR003957; CBRF_NF1B_topis.
DR InterPro; IPR001241; DNA_topoisolt.
DR InterPro; IPR002205; DNA_topoisolv.
DR Pfam; PF00204; DNA_gyraseb; 1.
DR Pfam; PF00521; DNA_topoisolv; 1.
DR Pfam; PF02518; HATPase_c; 1.
DR PRINTS; PR00615; CCAATSUBUNITA.
DR PRINTS; PR00418; TP12FAMILY.
DR ProDom; PD000742; DNA_topoisolv; 1.

```

```

DR SMART: SMO0387; HATPase C; 1.
DR SMART: SMO0433; TOP2C; 1.
DR SMART: SMO0434; TOP2C; 1.
DR PROSITE: PS00177; TOPOLISOMERASE_II; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
FT NP BIND 144 149
FT ACT SITE 830 830
FT DOMAIN 271 281
FT DOMAIN 308 316
FT DOMAIN 1089 1093
FT DOMAIN 1227 1234
FT SEQUENCE 1398 AA; 161029 MW; 5AAD7BE868FE5BE9 CRC64;
Query Match 11.0%; Score 99; DB 1; Length 1398;
Best Local Similarity 22.2%; Pred. No. 12;
Matches 44; Conservative 45; Mismatches .67; Indels 42; Gaps 9;
QY 1 KIIVKDDAR-----NTVKEFLINKDTGVESELKPHRVTVTIONGKMSSTIVSEE 51
Db 1093 KVLVEELRYGDPDYDKINIKKKEIEFGELDLDAAD-NPE-----DNEELIAGITVKDY 1145
QY 52 DFIL-DVYKGELEKGYQFGDGEWISGFEKKADGAYINLSKDTFIKVFKKIE----- 103
Db 1146 DYLLSMIFSLTLEK---VEDLLTQLKEKERETELRNITVETWMLDKIEKVEAIEFQR 1202
QY 104 -----KKEENKPFEDVSKKKDNQVNHSQLNESHREDLQREHSQ---KSDTKDYTA 155
Db 1203 NVELSNNEESNK-FVKARKQ-----GPSKKKKKKKKKKLSSDESGGDTSDSEFLVN 1255
QY 156 TVLDKNNISSKSTNNPN 173
Db 1256 TLNIKNTKNTTSSNN 1273
RESULT 3
YH00_YEAST STANDARD; PRT; 1345 AA.
ID YH00_YEAST
AC P38870;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 149.7 kDa protein in IRE1-KSP1 intergenic region.
GN YH080C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RN [1]
RS SEQUENCE FROM NP.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favellio A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Larrelle P., Louis E.J., Macri C., Maddis E., Meneses S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII";
RL Science, 265:2077-2082(1994).
CC -1- SIMILARITY: TO YEAST YFL042C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1eb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U10556; AAB68895.1; -.
CC PIR; S46817; S46817.
DR

```

```

DR SGP: S0001122; YHR080C.
DR InterPro: IPR004182; GRAM_dom.
DR Pfam: PF02893; GRAM_1.
DR SMART: SM00568; GRAM_1.
DR Hypothetical protein; Transmembrane.
DR TRANSMEM 1198 1218 POTENTIAL.
DR SEQUENCE 1345 AA; 149679 MW; 2FDAB94686564C2 CRC64;

Query Match 10.9%; Score 97.5; DB 1; Length 1345;
Best Local Similarity 27.0%; Pred. No. 15;
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY 59 KGELEKGVDFQDWEISGFEKCK-DAGVYVNLKDTFKVFKIEKKEENKPTFEDSK 117
DB 1109 KGATKKG-----SVGGQKVSVDYMLSEKRDII-----SRKSKKPKYKWK 1149
QY 118 KKDQNVHNSHLSHREKEDLQREHSHOKSDSTKVATVLD--KNNISKSTN---NP 172
DB 1150 SHDKRPPHSHKVE-----QKSESRKSDNDKDIILTHIDFVQNNFSEIFNMKLLSP 1201
QY 173 NK 174
DB 1202 QK 1203

RESULT 4
MAPB_HUMAN STANDARD; PRT; 2468 AA.
AC P46821;
DT 01-NOV-1995 (Rel. 32; Created)
DT 01-NOV-1995 (Rel. 32; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP 1 light chain LC1].
GN MAP1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95104835; PubMed=7806212;
RA Lien L.V., Reener C., Fischbach N., Kunkel L.M.;
RT "Cloning of human microtubule-associated protein 1B and the
RT identification of a related gene on chromosome 15."
RL Genomics 22:273-280(1994).
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1A by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (by similarity).
CC -1- SIMILARITY: TO MAP1A.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL, L06237; AAA18904.1; -

```

```

DR Genew: HGNC:6836; MAP1B.
DR MIM: 157129; -.
DR GO: GO:0005875; C:microtubule associated complex; TAS.
DR InterPro: IPR00102; MAP1B_neutrakin.
DR Pfam: PF00414; MAP1B_neutrakin; 10.
DR PROSITE: PS00230; MAP1B_NEURAXIN; 6.
DR Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2468
FT REPEAT 1878 1894 MAP1B 1.
FT REPEAT 1895 1911 MAP1B 2.
FT REPEAT 1912 1928 MAP1B 3.
FT REPEAT 1929 1945 MAP1B 4.
FT REPEAT 1946 1962 MAP1B 5.
FT REPEAT 1963 1979 MAP1B 6.
FT REPEAT 1997 2013 MAP1B 7.
FT REPEAT 2014 2030 MAP1B 8.
FT REPEAT 2031 2047 MAP1B 9.
FT REPEAT 2048 2064 MAP1B 10.
FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT KKEE AND KKEI/V REPEATS).
SQ SEQUENCE 2468 AA; 270618 MW; 540839C8DF09D461 CRC64;

Query Match 10.9%; Score 97.5; DB 1; Length 2468;
Best Local Similarity 24.8%; Pred. No. 29;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 14 KEPIANKDTGEVSELEKPRRTVTITONGKEMSTVSEDFILPYKGLKGYDFDWEI 73
DB 584 EKVMWKDKPKVTEKTPSVTEKEVPSKEPS-----PV-KAEVA-----EK 623
QY 74 SGFEKGDAGVYVNLKDTFKVFKIEKKEENKPTFVSKKQDQVHNSH 133
DB 624 QATDVKPPRAAEKRYKKEIKYKVP-----EDKKEKKEKKEVAKKEDTPI---KKEKP 675
QY 134 RKEDLORE-----EHSQKSDSTKV 153
DB 676 KKEVYKKEVKEIKKEEKEKPEKVEV 700

RESULT 5
GYRA_STABP STANDARD; PRT; 893 AA.
AC P54112;
DT 01-OCT-1996 (Rel. 34; Created)
DT 28-FEB-2003 (Rel. 41; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE DNA gyrase subunit A (Ec 5.99.1.3).
GN GYRA OR SE0005.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Sreedharan S., Peterson L.R., Fisher L.M.;
RA Chen Z., Wen Y., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN 121
RP SEQUENCE OF 1-94 FROM N.A., AND MUTAGENESIS OF SER-84.
RX MEDLINE=92102204; PubMed=1662027;
RA "Ciprofloxacin resistance in coagulase-positive and -negative
RA staphylococci: role of mutations at serine 84 in the DNA gyrase A
RT protein of Staphylococcus aureus and Staphylococcus epidermidis."
RL Antimicrob. Agents Chemother. 35:2151-2154(1991).
CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE

```



DR InterPro: IPR004868; DNA pol B\_2.  
 DR Pfam: PF03175; DNA\_pol\_B\_2; 1.  
 DR SMART: SM00486; POLB; 1.  
 DR PROSITE: PS00116; DNA\_POLYMERASE\_B; 1.  
 DR Transferrase; DNA-directed DNA polymerase; DNA replication;  
 DR DNA-binding; Plasmid; Mitochondrion.  
 KW DNA-binding; Plasmid; Mitochondrion.  
 SQ SEQUENCE 1202 AA; 138279 MW; 51D41FCBDBP2CDE CRC64;

Query Match 10.6%; Score 95; DB 1; Length 1202;

Best Local Similarity 22.8%; Pred. No. 20;  
 Matches 44; Conservative 37; Mismatches 80; Indels 32; Gaps 9;

QY 10 NTTFVEFLNK---DTGSESELKPHRVTVTTQNGKEMSTI---VSEEDF--ILPVKXG 60  
 DB 305 NTFQFFVYNAKIKRPTGNVRSIGFQNT-ITDKEITLITLALFLEREDITHTWSYDEG 363  
 QY 61 ELEK-----GYQPDGWEISGFEKKDAGYVNLKSDTFIKPKIE-----EKKEE 108  
 DB 364 DIDESEKPKGSLSPFQPKLKEGTQYANYTPPIKKDIYVQINKKINFGDLPKTMDL 423  
 QY 109 NK-PFFVDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKQVATVLDKNI----- 163  
 DB 424 SKPPLKLNKDKTSEIRMTIKNNQSYDI--IGHMINDENYITFRRAVDNSIIKIF 481  
 QY 164 ---SSKSTNNPN 173  
 DB 482 TVTDSMGNTNDPN 494

## RESULT 8

HS70\_PYRSA STANDARD; PRT; 649 AA.

AC P37899;  
 DT 01-OCT-1994 (Rel. 30; Created)  
 DT 01-OCT-1994 (Rel. 30; Last sequence update)  
 DT 28-FEB-2003 (Rel. 41; Last annotation update)  
 DE Heat shock 70 kDa protein.  
 GN HSP70.  
 OS Pyrenomonas salina.  
 OG Nucleomorph.  
 OC Eukaryota; Cryptophyta; Cryptomonadaceae; Pyrenomonas.  
 OX NCBI\_TaxID=3034;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94268506; Pubmed=8208251;  
 RA Hofmann C.J.B., Rensing S.A., Haebler M.M., Martin W.F., Mueller S.B.,  
 RA Couch J., McFadden G.I., Igloi G.L., Mäler U.-G.;  
 RT "The smallest known eukaryotic genomes encode a protein gene: towards  
 an understanding of nucleomorph functions."  
 RL Mol. Genet. 243:600-604(1994).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -2-  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X72621; CA51197.1; -  
 DR PIR: S42488; S42488.  
 DR HSSP: P08109; ICRK.  
 DR InterPro: IPR001023; HSP70.  
 DR Pfam: PF00012; HSP70; 1.  
 DR PRINTS: PR00301; HEATSHOCK70.  
 DR ProDom: PD000089; HSP70; 1.  
 DR PROSITE: PS00297; HSP70\_1; 1.  
 DR PROSITE: PS00329; HSP70\_2; 1.  
 DR PROSITE: PS01036; HSP70\_3; 1.  
 KW ATP-binding; Heat shock; Nucleomorph.  
 SQ SEQUENCE 649 AA; 72079 MW; B627B08FP90C9164 CRC64;

Query Match 10.4%; Score 93.5; DB 1; Length 649;  
 Best Local Similarity 24.5%; Pred. No. 13;  
 Matches 38; Conservative 26; Mismatches 50; Indels 41; Gaps 7;

QY 17 ILNKDTGVESELKPHRVTVTTQNG---KEMSTIVSEDFILPVKGELEKGYQPDGWEI 73  
 DB 491 ILNVASDPSKTSKSNKRTITNDKGLSKKEIERMVEAE-----KYKTEDEK----- 537  
 QY 74 SGFEKKDA-----GYVNLK---KDTFIKPKIEKEKEENKPTDVSKKKNPQVN 125  
 DB 538 -LDKKLEAKNSLENVANINVTVD-----EKLKKEIOEDKKSIEKYE----- 582  
 QY 126 HSQLNESHKEDLQREHSQKSDSTKQVATVLDK 160  
 DB 583 -VLEFPTNEDLEKEKEKEKEKLNANPLISK 615

## RESULT 9

ADSV\_HUMAN STANDARD; PRT; 715 AA.

AC Q9Y6U3; Q9Y6U3;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 28-FEB-2003 (Rel. 41; Last sequence update)  
 DT 15-SEP-2003 (Rel. 42; Last annotation update)  
 DE Adseverin (Scinderin).  
 GN SCIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=22388257; Pubmed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heisch F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Caeava T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Hellon E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Schnerker A., Schein J.E., Jones S.J.W., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC [2]  
 RP SEQUENCE OF 1-527 FROM N.A.  
 RA Kallicki J., Smith-Craig R.;  
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 248-715 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Itoigai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Akita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuo Y., Oshima A.;  
 RT "NERO human cDNA sequencing project";  
 RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: Ca(2+)-dependent actin filament-severing protein that is  
 presumed to have a regulatory function in exocytosis by affecting  
 the organization of the microfilament network underneath the  
 plasma membrane. In vitro, also has barbed end capping and  
 nucleating activities in the presence of Ca(2+).

CC -1- SIMILARITY: BELONGS TO THE VILILIN/GELSOLIN FAMILY.  
CC -1- SIMILARITY: Contains 6 gelsolin-like repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: BC021090; AAH21090.1; -  
CC EMBL: AC005281; AAD15423.1; -  
CC EMBL: AK027778; BAB55361.1; -  
CC HSSP: P02640; 2VIL.  
CC InterPro: IPR001974; Gelsolin.  
CC Pfam: PF00626; Gelsolin; 6.  
CC PRINTS: PR00597; GELSOLIN.  
CC SMART: SM00262; GEL; 6.  
CC Cytoskeleton; Actin-binding; Repeat: Calcium; Capping protein.  
CC ACTIN-SEVERING (POTENTIAL).  
CC CA(2+)-DEPENDENT ACTIN BINDING.  
CC DOMAIN 1 363  
CC FT REPEAT 364 715 GELSOLIN-LIKE 1.  
CC FT REPEAT 27 76 GELSOLIN-LIKE 1.  
CC FT REPEAT 148 188 GELSOLIN-LIKE 2.  
CC FT REPEAT 265 307 GELSOLIN-LIKE 3.  
CC FT REPEAT 398 451 GELSOLIN-LIKE 4.  
CC FT REPEAT 523 564 GELSOLIN-LIKE 5.  
CC FT REPEAT 626 668 GELSOLIN-LIKE 6.  
CC FT SITE 112 119 POLYPHOSPHOINOSITIDE BINDING (BY  
CC SIMILARITY).  
CC FT SITE 138 146 POLYPHOSPHOINOSITIDE BINDING (BY  
CC SIMILARITY).  
CC FT CONFLICT 61 61 R -> H (IN REF. 2).  
CC FT SEQUENCE 715 AA; 80508 MW; 45FBE42CBCEFDDB0 CRC64;  
SQ  
Query Match 10.4%; Score 93; DB 1; Length 715;  
Best Local Similarity 21.0%; Pred. No. 16;  
Matches 37; Conservative 39; Mismatches 58; Indels 42; Gaps 6;  
QY 3 VKKDFARNTTVEFLINKDTGEVSELPKRVVTYVTONGKEMSTIYSEDFILPYVKGEL 62  
DB 239 ITADISNRKAKLYWSDASGSM-----RVTVVAENEPFAMMLISECFILD--HGAA 290  
QY 63 EKGYPDGEIGSEIFSEK-----KDAGYVINKSCQFEIK-----PVFKK----- 100  
DB 291 KQIFWKGKQANPQEKRAAKTAEBFLQOMNYSKNTQIOVLPEGGETPLFKQFFKMDRK 350  
QY 101 -----IEKKEENKPTFDVSKKKDNQV--NHSQLNESHKEDLOREHS 144  
DB 351 DSDGEGKYVTEKVAQIKQIPFDASKLHSSPQMAAQHMVDDGSGKVAIMVENN 406  
RESULT 10  
MAPB MOUSE STANDARD; PRT; 2464 AA.  
ID MAPB MOUSE  
AC P14673;  
DT 01-APR-1990 (rel. 14, Created)  
DT 01-APR-1990 (rel. 14, Last sequence update)  
DT 28-FEB-2003 (rel. 41, Last annotation update)  
DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))  
DE [Contains: MAP1 light chain LC1].  
DE GN MAP1B OR MAP1B OR MAP1B.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND DOMAIN.  
RC STRAIN=Swiss Webster; TISSUE=Brain;  
RA MEDLINE=90094539; PubMed=2480963;  
RX Noble M., Lewis S.A., Cowan N.J.;  
RT "The microtubule binding domain of microtubule-associated protein  
RT MAP1B contains a repeated sequence motif unrelated to that of MAP2

RT and tau.";  
RL J. Cell Biol. 109:3367-3376(1989).  
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS  
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
CC STABILIZING MICROTUBULES.  
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
CC WITH MAP1A AND MAP1B PROTEINS.  
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
CC responsible for the binding of MAP1B to microtubules.  
CC -1- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED  
CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH  
CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION  
CC OF MAP1B.  
CC -1- SIMILARITY: TO MAP1A.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X51386; CAA35761.1; -  
CC PIR: S07549; QRMSP1.  
CC MGD; MG1:1306778; Metap1b.  
CC GO; GO:0016358; P: dendrite morphogenesis; IMP.  
CC GO; GO:0001578; P: microtubule bundling; IMP.  
CC InterPro: IPR00102; MAP1B neuraxin.  
CC Pfam: PF00414; MAP1B neuraxin; 10.  
CC PROSITE: PS00230; MAP1B\_NEURAXIN; 7.  
CC KW Microtubules; Repeat; Phosphorylation.  
CC FT CHAIN ? 2464  
CC FT REPEAT 1874 1890 MAP1B 1.  
CC FT REPEAT 1891 1907 MAP1B 2.  
CC FT REPEAT 1908 1924 MAP1B 3.  
CC FT REPEAT 1925 1941 MAP1B 4.  
CC FT REPEAT 1942 1958 MAP1B 5.  
CC FT REPEAT 1959 1975 MAP1B 6.  
CC FT REPEAT 1993 2009 MAP1B 7.  
CC FT REPEAT 2010 2026 MAP1B 8.  
CC FT REPEAT 2027 2043 MAP1B 9.  
CC FT REPEAT 2044 2060 MAP1B 10.  
CC FT DOMAIN 589 787  
SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDDBA87 CRC64;  
Query Match 10.4%; Score 93; DB 1; Length 2464;  
Best Local Similarity 27.8%; Pred. No. 61;  
Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;  
QY 14 KEFLINKDTGEVSELPKRVVTYVTONGKEMSTIYSEDFILPYVKGELKGYQFDGWEI 73  
DB 584 EKVLYKKDQPVYTESKP--SVT---EKVSS---KEE--SPV-KAEVA-----EK 623  
QY 74 SGEFGKQAGYVINKDTFIKRVK-KLEEKEENKPTFDVSKKKDNQVNHSQLNES 132  
DB 624 QATESKP-----KTVKQVAKKEIKTKLEKKEE--KPKKVVKKEDTPTL---KDEK 672  
QY 133 HKKEDLORE-----EISQSDSTKQVATATYLDGNINSSKSTNNPNK 174  
DB 673 PKKEVYKKEIKKEIKERKELKEVKEKTPDKAKKEVKEEKEKKEKPEPK 728  
RESULT 11  
DACA BACSU STANDARD; PRT; 443 AA.  
ID DACA BACSU  
AC P08750;  
DT 01-AUG-1988 (rel. 08, Created)



DT 01-OCT-1994 (Rel. 30, last sequence update)  
 DT 28-FEB-2003 (Rel. 41, last annotation update)  
 DE D-alanyl-D-alanine carboxypeptidase precursor (EC 3.4.16.4) (DD-  
 DE peptidase) (DD-carboxypeptidase) (CPase) (PAPS).  
 GN DACC.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 CX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96051385; PubMed=7584024;  
 RA Ogasawara N., Nakai S., Yoshikawa H.;  
 RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
 RT subtilis chromosome containing the replication origin.";  
 RL DNA Res. 1:1-14(1994).  
 (2)  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96044033; PubMed=9384377;  
 RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Berteiro M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Dentzer F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabel C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Jorts B., Karamata D., Kasahara Y., Klaerr-Blanchard N., Klein C.,  
 RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kunita K., Lapidus A., Lardinois S., Lamber J., Lazarovic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Priebean E., Puje P., Fumelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Setguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarelli A.,  
 RA Viart A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yaman K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumslein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL Nature 390:249-256(1997).  
 (3)  
 RN [3]  
 RP SEQUENCE OF 32-102.  
 RX MEDLINE=80182289; PubMed=6768745;  
 RA Waxman D.J., Strominger J.L.;  
 RT "Sequence of active site peptides from the penicillin-sensitive D-  
 RT alanine carboxypeptidase of Bacillus subtilis. Mechanism of  
 RT penicillin action and sequence homology to beta-lactamases.";  
 RL J. Biol. Chem. 255:3964-3976(1980).  
 (4)  
 RN [4]  
 RP SEQUENCE OF 103-443 FROM N.A.  
 RX MEDLINE=86250602; PubMed=3087956;  
 RA Todd J.A., Roberts A.N., Johnstone K., Piggett P.J., Winter G.,  
 RA Ellar D.J.;  
 RT "Reduced heat resistance of mutant spores after cloning and  
 RT mutagenesis of the Bacillus subtilis gene encoding penicillin-binding  
 RT protein 5.";  
 RL J. Bacteriol. 167:257-264(1986).  
 RN [5]  
 RP SEQUENCE OF 414-443.  
 RX MEDLINE=81117303; PubMed=6780559;  
 RA Waxman D.J., Strominger J.L.;  
 RT "Primary structure of the COOH-terminal membranous segment of a

RT penicillin-sensitive enzyme purified from two Bacilli.";  
 RL J. Biol. Chem. 256:2067-2077(1981).  
 CC -1- FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE  
 CC CELL WALL PRECURSORS.  
 CC -1- CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O = 2 D-alanine.  
 CC -1- PATHWAY: Peptidoglycan synthesis; final stages.  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11; ALSO KNOWN AS THE  
 CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: D26185; BAA05246.1; -;  
 CC DR EMBL: Z99104; CAB11786.1; -;  
 CC DR EMBL: M13766; AAA22375.1; -;  
 CC DR PIR: S66040; S66040.  
 CC DR MEROPS: S11.001; -;  
 CC DR Subtilist; BG10074; dacc.  
 CC DR InterPro: IPR001967; Ala/AlaCBptase1.  
 CC DR Pfam: PF00768; Peptidase\_S11; 1.  
 CC DR PRINTS: PR00725; DA0CBPTASE1.  
 CC KW Hydrolysis; Carboxypeptidase; Peptidoglycan synthesis; Cell wall;  
 CC Membrane; Signal; Complete proteome.  
 CC FT CHAIN 1 31  
 CC FT ACT SITE 32 443 D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.  
 CC FT CONFLICT 100 100 E -> Q (IN REF. 3).  
 CC FT CONFLICT 227 227 E -> Q (IN REF. 4).  
 CC SQ SEQUENCE 443 AA; 48636 MW; D46C5B0307D7C117 CRC64;  
 CC -----  
 CC Query Match 10.3%; Score 92.5; DB 1; Length 443;  
 CC Best Local Similarity 25.0%; Pred. No. 10; Indels 33; Gaps 6;  
 CC Matches 36; Conservative 26; Mismatches 49;  
 CC QY 6 DPA-RNTTVKEPFLNKDTGEVSELEKPHRVTVTQNGKMSSTVSEEDPILPVYKGELEK 64  
 CC DB 304 DYAPNPFMKIYAKGQ-----QYKGRK-TISVDKGRKEKEGIVTNKAFSLPVKNGE- EK 356  
 CC QY 65 GYQPDGWEISGPEGRKQAGVYINISKDTFIKPVFKIIEKKBEKKPTFDVSKKQNDPOV 124  
 CC DB 357 NYKAK-----VTINKDNLTPVKKGTGVK-----LTAEYTGDEK 391  
 CC QY 125 NKSQNESHRKEDLQREHSQKSD 148  
 CC DB 392 DYGLNSDLAAGVLTAKENVERKAN 415  
 CC -----  
 CC RESULT 12  
 CC IGA2\_HAEMIN STANDARD; PRT; 1702 AA.  
 CC AC P45384;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, last annotation update)  
 CC DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).  
 CC GN IGA.  
 CC OS Haemophilus influenzae.  
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 CC OC Pasteurellaceae; Haemophilus.  
 CC CX NCBI\_TaxID=727;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=HK715 / Serotype B;  
 CC RX MEDLINE=92234949; PubMed=1373717;  
 CC RA Poulsen K., Reinholdt J., Kilian M.;  
 CC RT "A comparative genetic study of serologically distinct Haemophilus  
 CC influenzae type 1 immunoglobulin A1 proteases.";

```

RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTERACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M87489; AAA24966.1; -
CC PIR; A41859; A41859.
CC MEROPS; S06.001; -.
CC InterPro: IPR006315; Autotransport.
CC InterPro: IPR005546; Autotransporter.
CC InterPro: IPR000710; IGA_S6.
CC InterPro: IPR004899; Peptactin.
CC Pfam; PF03797; Autotransporter; 1.
CC Pfam; PF02395; IGA1; 1.
CC Pfam; PF03212; Peptactin; 1.
CC PRINTS; PR00921; IGASERPRASE.
CC TIGRfam; TIGR01414; autotrans_bar1; 1.
CC HydroLase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
CC SIGNAL 1 25
CC CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
CC PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
CC ACT SITE 288 288 PROBABLE.
CC DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
CC K.
CC REPEAT 1109 1116 1.
CC REPEAT 1117 1124 2.
CC FT REPEAT 1109 1116 1.
CC FT REPEAT 1117 1124 2.
CC SQ SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

Query Match 10.3%; Score 92.5; DB 1; Length 1702;
Best Local Similarity 23.9%; Pred. No. 44;
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;

QY 2 IYVKDFARNT---TYKEFLINKDGEVSELEKPHRVTVTIQNGKEMSTIVSE---EDPL 54
DB 1206 VVSKNOTENTTTDPTREKTAKVETKTQE--PPVVASQASPKQOSEFVQPOAVLESEN 1263
QY 55 LPVYKGELEKGYQFDGMEISGFEKQDA-GYVINLSKDTPIKPVFKIEKEENKPTF 113
DB 1264 VFTVNAEVEVQQLQVTSATVSTKQPAPENSINTSATAITETAKSKPKQETASTAE 1323
QY 114 DYSKKKDNPOVNHSQLNESHREKDLQREHS---QKSDTKDYTAIVLDKNNISSKSTN 170
DB 1324 DASQHKANTVADNSVANNSESESPKRRRRRISIQPQETSAEETTAASTDETTIADNSKS 1383
QY 171 NPKK 174
DB 1384 KPRK 1387

RESULT 13
YDT2 SCHPO STANDARD; PRT; 1888 AA.
ID YDT2 SCHPO
AC 014207;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE Hypothetical protein C6B12.02c in chromosome I.
GN SPAC6B12.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972.
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Pat N., Hayles S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jags J.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean U.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor J., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Gymnopoulos B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinhardt R., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wandt R., Purnelle B.,
RA Goiffreau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Moutier S.,
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen O., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
RA "The genome sequence of Schizosaccharomyces pombe."
RT Nature 415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z98531; CAB11064.1; -.
CC PIR; T39009; T39009.
CC DR GeneDB SPombe; SPAC6B12.02c; -.
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 697 717 POTENTIAL.
CC FT TRANSMEM 755 775 POTENTIAL.
CC FT TRANSMEM 866 886 POTENTIAL.
CC FT TRANSMEM 915 935 POTENTIAL.
CC FT TRANSMEM 977 997 POTENTIAL.
CC FT TRANSMEM 1033 1053 POTENTIAL.
CC FT TRANSMEM 1336 1356 POTENTIAL.
CC FT TRANSMEM 1645 1665 POTENTIAL.
CC FT DOMAIN 1662 1665 POLY-LEU.
CC SQ SEQUENCE 1888 AA; 217432 MW; 8AD3BBCE32397C29 CRC64;

Query Match 10.3%; Score 92.5; DB 1; Length 1888;
Best Local Similarity 23.1%; Pred. No. 49;
Matches 39; Conservative 25; Mismatches 76; Indels 29; Gaps 5;

QY 11 TTYKEFLINKDGEVSELEKPHRVTVTIQNGKEMSTIVSE---EDPLPYVYKGELEKGYQFD 69
DB 394 TSENPFQNVANAAVSTIPVATTTTKMKQRKFVVEVEXKLPDLLESEY----- 442
QY 70 GWEISGEGSKDAGVINLSKDTPIKPVFKIEKEENKPTFVYSKKKDNPOVNHSQL 129
DB 443 -----GKKAPKFLRVAFARSSSHIP--KMIIRKROMDSKRYFSDKESDRQVLDVLS 492

```

QY 130 NESHKEDLOREHSOKSDS-TRDYATATVLDKN-----NISKSTNN 171  
 DB 493 DWYSGKHELVOGSHSYKPKSDSKSVGNIFSVNSKSHNINAKTAPANN 541

RESULT 14

NSBL\_MOUSE STANDARD; PRT; 406 AA.  
 AC Q9ULJ35; O88832; Q8VC71; Q9CUM1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Nucleosome binding protein 1 (Nucleosome binding protein 45) (NBP-45) (GARP45 protein).  
 GN NSBP1 OR GARP45.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;

[1]  
 SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
 RX MEDLINE=20158948; PubMed=10692437;  
 RA Shitakawa H., Landsman D., Postnikov V.V., Bustin M.;  
 RT "NBP-45, a novel nucleosomal binding protein with a tissue-specific and developmentally regulated expression.";  
 RL J. Biol. Chem. 275:6368-6374(2000).

[2]  
 SEQUENCE FROM N.A.  
 RA Onoda G., Suzuki N., Saito H., Honda T., Sato H., Kuwano R.;  
 RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.

[3]  
 SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stopleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ueda T.B., Toshitsuki S., Carrinci P., Prange C., Rana S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosnak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalski U., Smallus D.E., Schnerch A., Schin J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[4]  
 SEQUENCE OF 7-195 FROM N.A.  
 RP STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Iwama H., Nishi K., Kiyosawa H., Kondo S., Yamakura I., Saito T., Okazaki Y., Gojohori T., Bono H., Kaubukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schriml L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M., Guttingich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momdearts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilting L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kontani S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).

CC - FUNCTION: Binds specifically to nuclear core particles and acts as  
 a transcriptional activator.  
 CC - SUBCELLULAR LOCATION: Nuclear.  
 CC - TISSUE SPECIFICITY: Widely expressed with highest levels in  
 submaxillary gland, thymus, kidney and liver and lowest levels in  
 brain, lung, pancreas and eye.  
 CC - DEVELOPMENTAL STAGE: Highest levels are found in 7-day-old  
 embryos. Levels in the 7-day-old embryo are 4-fold higher than in  
 the adult and almost 10-fold higher than in later embryonic  
 stages.

CC - SIMILARITY: BELONGS TO THE HMG-14/HMG-17 PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; AF213454; AAF30179.1; -  
 DR EMBL; AB018374; BAA33783.2; -  
 DR EMBL; BC021626; AAB21626.1; -  
 DR EMBL; AK013748; BAB28982.1; -  
 DR MGD; MGI:1355295; Nsdp1.  
 DR GO; GO:0005654; C:nucleoplasm; IDA.  
 DR GO; GO:0003682; P:chromatin binding activity; IDA.  
 DR GO; GO:0006356; P:regulation of transcription from Pol I prom. .; IDA.  
 DR InterPro: IPR000079; HMG\_14\_17.  
 DR Pfam: PF01101; HMG14\_17\_1.  
 DR SMART; SM00527; HMG17; 1.  
 DR PROSITE; PS00355; HMG14\_17; FALSE\_NEG.  
 KW Transcription regulation; Activator; DNA-binding; Nuclear protein.  
 FT CONFLICT 74 74 M -> V (IN REF. 1).  
 FT CONFLICT 390 390 N -> H (IN REF. 2).  
 SQ SEQUENCE 406 AA; 45344 MW; 59A4305613EC9679 CMC64;

Query Match 10.3%; Score 92; DB 1; Length 406;  
 Best Local Similarity 27.4%; Pred. No. 10;  
 Matches 48; Conservative 20; Mismatches 59; Indels 48; Gaps 9;

QY 1 KIVKDPKRNNTVVEFLINDTGEVSELRPRVVTIONGK-----EMSTVSEEDF 53  
 DB 126 KANIQDVKEDDGE--HKDTG-----EVDGKIEEGKNEKGTAKSD- 169

QY 54 ILPVYKGLKGYQFDGWEISGFEGKDGAVINLSKD-----TFIKPVFKLIEK 104  
 DB 170 -AEVSKDEEKGDNK-----EDGKEGEGKEEKKDCKGDTGTEVAGEQKNEAED 222

QY 105 ---KEEENKPTFDVSKKKNPQVNHSQLNESHKREDLQREE--HSQKSDSTKOV 153  
 DB 223 DGKCKEENK---EVGKEGQPEEDGKEDLHREVGKEDLHREDGKEGQPEEDGKEI 274

RESULT 15

SPT7 YEAST STANDARD; PRT; 1332 AA.  
 ID SPT7 YEAST  
 AC P35177;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Transcriptional activator SPT7.  
 GN SPT7 OR YBR081C OR YBR0739.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OK NCBI\_TaxID=4932;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=8288C;  
RA MEDLINE=95229044; PubMed=7713415;  
RX Ganehroff L.J., Dollard C., Tan P., Winston F.;  
RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein  
important for transcription in vivo.";  
RL Genetics 139:523-536(1995).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=8288C;  
RA MEDLINE=95076715; PubMed=7985423;  
RX van der Aart O.J.M., Barthe C., Doignon F., Aigle M., Crouzet M.,  
RA Steensma H.Y.;  
RT "Sequence analysis of a 31 kb DNA fragment from the right arm of  
Saccharomyces cerevisiae chromosome II.";  
RL Yeast 10:959-964(1994).  
RN [3]  
RN SEQUENCE OF 1-835 FROM N.A.  
RA STRAIN=8288C;  
RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestearazu A.,  
RA Vissers S.;  
RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RN SEQUENCE OF 463-523 FROM N.A.  
RX MEDLINE=92285152; PubMed=1350857;  
RA Haynes S.R., Dollard C., Winston F., Beck S., Trowdale J.,  
RA David I.B.;  
RT "The bromodomain: a conserved sequence found in human, Drosophila and  
yeast proteins.";  
RL Nucleic Acids Res. 20:2603-2603(1992).  
RN [5]  
RN FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY  
OTHER GENES.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: Contains 1 bromodomain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, L22537; AAC37424.1; -  
DR EMBL, X76294; CAAS3940.1; -  
DR EMBL, Z35950; CAAS5026.1; -  
DR EMBL, M87651; AAA35087.1; -  
DR PIR, S41552; S41552.  
DR HSSP, Q92831; 1B91.  
DR TRANSFAC, T04835; -  
DR SGD, S0000285; SPT7.  
DR GO, GO:000124; C:SAGA complex; IDA.  
DR InterPro, IPR001487; Bromodomain.  
DR Pfam, PF00439; bromodomain; 1.  
DR PRINTS, PR00297; BROMODOMAIN.  
DR SMART, SM00297; BROMO.1.  
DR PROSITE, PS00633; BROMODOMAIN\_1; 1.  
DR PROSITE, PS50014; BROMODOMAIN\_2; 1.  
DR Transcription regulation; Nuclear protein; Activator; Bromodomain.  
FT DOMAIN 458 528 BROMODOMAIN.  
SQ SEQUENCE 1332 AA; 152616 MW; 083B6362466244F CRC64;

Query Match 10.3%; Score 92; DB 1; Length 1332;  
Best Local Similarity 21.7%; Pred. No. 37;  
Matches 50; Conservative 32; Mismatches 72; Indels 76; Gaps 12;

QY 1 KIVVDFPANTYK-----EPI-----LNKDTREVELKPHRV----- 33  
DB 480 QIIKSMIDLTVYKLUKLSFOYDSKQFVDIMLIMWNCCLTYSDPSHF--LKGHALAMOK 537  
QY 34 -----TVIINGKEMSTIVSEEDFILPVYKGELEKGYQFGWEISGPEGKDGAG 83

DB 538 KSLQILRIMPIITIRNADLEKEI---EDM-----EKDXYEDDEEEVAGSRKG-- 585  
QY 84 VYINLSKDTFK--PVKRIEKKKEENKPTFD-----VSKKD-----NP 122  
DB 586 --LNNGAHMLAENGKSVKSSKTKVQDAPFNDKLTSLVPEGEKDKTASSTVYVHE 643  
QY 123 QVNHQNLNESHREKDLQR-EESHQKSDSTKQVATVLD-KNNISKSTTN 170  
DB 644 NVNKNIEKNGKNEEQDWEESSTKEDSSKADAKKDTEDGLQKXTAEN 693

RESULT 16  
ID LIP STAMP STANDARD; PRT; 688 AA.  
AC 002510;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Lipase precursor (EC 3.1.1.3) (Glycerol ester hydrolase).  
GN GEHC OR SE0281.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RN SEQUENCE FROM N.A., AND SEQUENCE OF 303-315.  
RA STRAIN=9;  
RX MEDLINE=93171870; PubMed=8436947;  
RA Farrell A.M., Foster T.O., Holland K.T.;  
RT "Molecular analysis and expression of the lipase of Staphylococcus  
epidermidis.";  
RL J. Gen. Microbiol. 139:267-277(1993).  
RN [2]  
RN SEQUENCE FROM N.A.  
RA STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
RA Chen Z., Wen Y.;  
RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
fatty acid anion.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS  
NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE  
CONVERSION).  
CC -1- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, M95577; AAA19729.1; -  
DR EMBL, AE016744; AAA03878.1; -  
DR PIR, A47705; A47705.  
DR InterPro, IPR005877; Gpos\_Y5IRK.  
DR InterPro, IPR000734; Lipase.  
DR InterPro, IPR000379; Ser\_estr\_site.  
DR Pfam, PF04650; Y5IRK\_signal; 1.  
DR TIGRFAMs, TIGR01168; Y5IRK\_signal; 1.  
DR PROSITE, PS00120; LIPASE\_SER; 1.  
KW Hydrolase; Lipid degradation; Zymogen; Signal; Complete proteome.  
FT SIGNAL 1 35  
FT PROPEP 36 302  
FT CHAIN 303 688  
FT ACT\_SITE 418 418  
FT ACT\_SITE 648 648  
FT CONFLICT 96 96  
FT CONFLICT 120 120  
SQ SEQUENCE 688 AA; 77343 MW; 6C95DB3A7BAF66F CRC64;

Query Match 10.0%; Score 90; DB 1; Length 688;  
 Best Local Similarity 24.4%; Pred. No. 25;  
 Matches 39; Conservative 30; Mismatches 65; Indels 26; Gaps 7;

QY 19 NKDTGEVSELKPHRYVTIÖNGKEMSTIVSEDFILPYKGLKGYQPDGWEISGFE-77  
 DB 68 NQNVNEKSVN-----SITENESLHNERPKNEDW-----QQQKDSQNDKKSSEVVGQ 115

QY 78 GKKGAGYINLSKDTFIKPVFKKIE-EKKEENKPTFDVSKKKDNPQV--NHSQINESHR 134  
 DB 116 NKENAFVQNHSE--KPOQOELEKHAENNQTLHSAQSNEDVTKPSQDLNTAA 172

QY 135 KEDLQREHSQKSDSTKDTATVLDKNNISKSSTNNPK 174  
 DB 173 KOEDSOKENLSKDTQSSKTTDL-----RATANQSK 205

RESULT 17  
 YFK8 YEAST STANDARD; PRT; 778 AA.

AC P43610;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical 88.7 kDa helicase in CDC26-SAP155 intergenic region.  
 GN YFR038W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=95400292; PubMed=7670463;  
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,  
 RA Yamazaki M., Tashiro H., Eki T.,  
 RT "Analysis of the nucleotide sequence of chromosome VI from  
 RT Saccharomyces cerevisiae.";  
 RL Nat. Genet. 10:261-268(1995).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=96287654; PubMed=8666381;  
 RA Eki T., Naitou M., Hagiwara H., Abe M., Ozawa M., Sasanuma S.-I.,  
 RA Sasanuma M., Tsuchiya Y., Shibata T., Watanabe K., Ono A.,  
 RA Yamazaki M.-A., Tashiro H., Hanaoka F., Murakami Y.,  
 RT "Fifteen open reading frames in a 30.8 kb region of the right arm of  
 RT chromosome VI from Saccharomyces cerevisiae.";  
 RL Yeast 12:177-190(1996).  
 RL [2]  
 RP -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D50617; BAA09277.1; -  
 DR PIR: S56293; S56293.  
 DR SGD: S0001934; YFR038W.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR000330; SNF2\_N.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00176; SNF2\_N; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00490; HELIC; 1.  
 KW Hypothetical protein; Nuclear protein; DNA-binding; Helicase;  
 KW ATP-binding.

FT NP BIND 247 254 ATP (POTENTIAL).  
 FT SITE 352 355 DEGH BOX.  
 SQ SEQUENCE 778 AA; 88730 MW; 3B6C0857B5EABD84 CRC64;

Query Match 10.0%; Score 90; DB 1; Length 778;  
 Best Local Similarity 23.0%; Pred. No. 28;  
 Matches 44; Conservative 29; Mismatches 76; Indels 42; Gaps 6;

QY 5 KDFANNTVKEPILNKDTGEVSELKPHRYVTIÖNGKEMST-----IYSEDFILP 56  
 DB 26 RMSGVNVVREKEVNDLTADISDSDDSDSDNKHGKDNPATPWLQDVAHSDBI--- 82

QY 57 VKKGLKGYQPDGWEISGFEKKGAGYINLSKDTFIKPVFKKIEKKEENKPTFDVS 116  
 DB 83 -----QDSDSDSDTEAVQ-AQYVDKLAQDT--KSEOKSLDDEISEMDTIVSLK 129

QY 117 KKKDNPQVNHSQLNES-----HKREDLQR-----EHSQKSDSTKDTATATVL 158  
 DB 130 LKKLNEFVQSQVYSIIADTLHRSNRYANANTKDNNSDDEHSSKKRKTKKKSITDF 189

QY 159 DKNISKSSTT 169  
 DB 190 FKKQKKNEDTT 200

RESULT 18  
 TIG STAM STANDARD; PRT; 433 AA.

AC Q99T16;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trigger factor (TF).  
 GN TIG OR SAV1675 OR SA1499 OR MM1619.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699),  
 OS Staphylococcus aureus (strain N315), and  
 OS Staphylococcus aureus (strain MM2).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxId=158878, 158879, 196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mu50 / ATCC 700699, and N315;  
 RX MEDLINE=2111952; PubMed=11418146;  
 RA Kuroda M., Ono T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lin J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 RT aureus.";  
 RL Lancet 357:1225-1240(2001).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MM2;  
 RX MEDLINE=2240717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwana N., Asano K., Naito T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiramatsu K.,  
 RT "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA.";  
 RL Lancet 359:1819-1827(2002).  
 CC -1- FUNCTION: Involved in protein export. Acts as a chaperone by  
 CC maintaining the newly synthesized protein in an open conformation  
 CC (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY. TIG SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003363; BAB57837.1; -
DR EMBL; AP003134; BAB42766.1; -
DR EMBL; AP004827; BAB95484.1; -
DR PIR; A89951; A89951.
DR HAMAP; MF_00303; -.
DR InterPro; IPR001179; FKBP_PPIase.
DR InterPro; IPR005215; Trig_fac.
DR Pfam; PF00254; FKBP; 1.
DR TIGRFAMs; TIGR00115; trig; 1.
DR PROSITE; PS00453; FKBP_PPIASE_1; FALSE_NEG.
DR PROSITE; PS00454; FKBP_PPIASE_2; FALSE_NEG.
DR PROSITE; PS00059; FKBP_PPIASE_3; 1.
DR Cell division; Chapterone; Isomerase; Rotamase; Complete proteome.
FW FT DOMAIN 163 248 PPIASE, FKBP-TYPE.
SQ SEQUENCE 433 AA; 48609 MW; 8865D9AF6A1BC1E7 CRC64;

Query Match 10.0%; Score 89.5; DB 1; Length 433;
Best Local Similarity 22.1%; Pred. No. 16;
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 21 DTGEVSELSKPRVWT-T-IONGKE--MSSTIYSEEDFLIPYKKG-ELEK----- 64
Dd | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 81 DETDIKPAQPEVSVTQIEKGDPIFEATVVEPVPKIGDYKGLIEHQEFLSDDEIQE 140
QY 65 -----GYQPDGWEISG--FEGKDDAGVILNSKDTFIKP 96
Dd 141 AIDHSLGHLEWVWVEDGVENGDTVINIDFSG-SVDGEFEGQAGVLDLEIGSGSFI-P 198
QY 97 VFK-----KIEBKE-----EE--NKTEPVS---KKKNPQVNSQLNE- 131
Dd | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 199 GFEEQLBGMKVDDEEKDVVTFPEEYHAELAGKEATPFKYNVEIKFKXVEPLTDEIANEL 258
QY 132 -----SHRKEDLOREHESGKSDSTKDVTAIVLDKNNISKSTTN 170
Dd 259 DAEANTVDEYKENLKRKLAEQKATDAENV-----EKEEATITATDN 299

RESULT 19
PIK1_YEAST
ID PIK1_YEAST STANDARD; PRT; 1066 AA.
AC P39104;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Phosphatidylinositol 4-kinase PIK1 (BC 2.7.1.67) (PI4-kinase)
DE (Pcdins-4-kinase).
GN PIK1 OR YNL267W OR N0795.
OS Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JK9-3D;
RX MEDLINE=94425232; PubMed=8194527;
RA Garcia-Bustos J.F., Marini F., Stevenson I., Frei C., Hall M.N.;
RT "PIK1, an essential phosphatidylinositol 4-kinase associated with the
RL yeast nucleus.";
RN EMBL J. 13:2352-2361(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94069320; PubMed=8248783;
RA Flanagan C.A., Schnieders E.A., Emerick A.W., Kunisawa R., Admon A.,
RT "Phosphatidylinositol 4-kinase: gene structure and requirement for
RL yeast cell viability.";
RN Science 262:1444-1448(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SS288C / FY1679;

```

```
RA MEDLINE=96310631; PubMed=8740425;
RX Ser-Gupta M., Lyck R., Pleijg U.S., Niedenthal R.K., Hegemann J.H.;
RT "The sequence of a 24,152 bp segment from the left arm of chromosome
RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
RT genes."
RL Yeast 12:505-514(1996).
RC
CC -1- FUNCTION: ACTS ON PHOSPHATIDYLINOSITOL (PI) IN THE FIRST
CC COMMITTED STEP IN THE PRODUCTION OF THE SECOND MESSENGER
CC INOSITOL-1,4,5-TRISPHOSPHATE. PIK1 IS PART OF A NUCLEAR
CC PHOSPHOINOSITIDE CYCLE AND COULD CONTROL CYTOKINESIS THROUGH
CC THE ACTIN CYTOSKELETON.
CC CATALYTIC ACTIVITY: ATP + 1-phosphatidy1-D-myo-inositol = ADP +
CC 1-phosphatidyl-1D-myo-inositol 4-phosphate.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L76058; CA453658.1; -
CC DR EMBL; X20220; AAA34873.1; -
CC DR EMBL; X92494; CA463231.1; -
CC DR EMBL; Z71543; CA96174.1; -
CC DR PIR; A49335; AA9335.
CC SCD; S0005211; PIK1.
CC DR GO; GO:0005634; C:nucleus; IDA.
CC DR GO; GO:0016288; P:cytokinesis; IMP.
CC DR GO; GO:0006892; P:post Golgi transport; IGI.
CC DR InterPro; IPR000403; P13_P14_kinase.
CC DR Pfam; PF00454; P13_P14_kinase; 1.
CC DR SMART; SM00146; P13KC; 1.
CC DR PROSITE; PS00915; P13_4_KINASE_1; 1.
CC DR PROSITE; PS00916; P13_4_KINASE_2; 1.
CC DR PROSITE; PS50290; P13_4_KINASE_3; 1.
CC KW Transferase; Kinase; Nuclear protein.
CC FT DOMAIN 793 1041 P13X/P14K.
SQ SEQUENCE 1066 AA; 119922 MW; 7666979CA14B1CB5 CRC64;

Query Match 10.0%; Score 89.5; DB 1; Length 1066;
Best Local Similarity 22.0%; Pred. No. 43;
Matches 39; Conservative 34; Mismatches 59; Indels 45; Gaps 7;

QY 4 VKDFARTTYKVEFLINKDTGEVSELKHRYAVTVIQTIONKEMSSSTIVSEDFILPYKGELE 63
   |||::||::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 202 MKDPTKNMILLNKTLNKKTSRSKRVSNNR-----SSTPTSIDLDLP----- 244

QY 64 KGVOFDGMEISPEFGKKAGAYINLSKDTEFKPVFKKIEEKKEE-ENKPTFVDSKKKDN 121
   ::||::||::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 245 -----KTKEDASF--RKSRHSEVKLDFDIVDIGNQVFEEIRISSIKLPEKRK 289

QY 122 PQ-VNHSQLAESHRKEPDLOREHSQKSDSKVDYAIVLDKN---ISKSTTNPNPK 174
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 290 PKYILDNSYVHRYTDGKNINR-----DGSIISNTAKALDGNKGDIYSKGRDENNE 339

RESULT 20
IGAO_HAEIN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA OR IGA1 OR HI0990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
```

RP SEQUENCE FROM N.A.  
RC STRAIN-Serotype D;  
RA Wright A., Fishman Y., Tai F., Plaut A.G.;  
RL Submitted (May-1991) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RD / KM20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Felschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
McKenney K., Sutton R., Fitzhugh W., Fields C.A., Gocayne J.D.,  
Scott J.D., Shiley R., Liu L.-I., Glodde A., Kelley J.M.,  
Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
Usterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,  
Pine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,  
Rahm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
Venter J.C.;  
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd";  
RL Science 269:496-512(1995).  
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
CC PRODUCING INTRACT FC AND FAB FRAGMENTS.  
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule  
CC substrates are known.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X59800; -; NOT ANNOTATED\_CDS.  
DR EMBL; U32779; AAC22651.1; -;  
DR PIR; H64106; H64106.  
DR MEROPS; S06.001; -;  
DR TIGR; H10990; -;  
DR InterPro; IPR006315; Autotransporter.  
DR InterPro; IPR005546; Autotransporter.  
DR InterPro; IPR000710; IGA\_S6.  
DR InterPro; IPR004899; Pertactin.  
DR Pfam; PF03197; Autotransporter; 1.  
DR Pfam; PF02395; IGA1; 1.  
DR Pfam; PF03212; Pertactin; 1.  
DR PRINTS; PR00921; IGASERPTAS.  
DR TIGRFPfam; TIGR01414; autotransp\_bar1; 1.  
KW Hydrolyase; Serine protease; Transmembrane; Zymogen; signal;  
KW Complete proteome.  
FT SIGNAL 1 25  
FT CHAIN 26 1014  
FT PROPEP 1015 1694  
FT ACT\_SITE 288  
FT CONFLICT 253 254  
FT CONFLICT 272 272  
FT CONFLICT 464 464  
FT CONFLICT 866 866  
FT CONFLICT 1036 1036  
FT CONFLICT 1074 1074  
FT CONFLICT 1421 1421  
FT CONFLICT 1545 1545  
SQ SEQUENCE 1694 AA; 185539 MW; CS2427013F93178C CRC64;  
Query Match 10.0%; Score 89.5; DB 1; Length 1694;

Best Local Similarity 23.4%; Pred. No. 71;  
Matches 43; Conservative 27; Mismatches 101; Indels 13; Gaps 5;  
QY 2 IVWDPAFNT-----TWSEFLNKDTGEVSELKPHRYVTYITONGEMSSITYS---EDPI 54  
DB 1198 VVSXKQNTNTDOPFEREKTKVETKQ---PPVVSQAQPKQOSTVOPQAVLESEN 1255  
QY 55 LPVYKGELEKGYQDFGWEISGFEGKDA-GVYVINSKOTFIKPYFKIIEKKEENKPTF 113  
DB 1256 VPTVNAAEVQAQLOTQTSATVSTKQAPENSINTGATALTERRKSDKQETIATASTE 1315  
QY 114 DVSKKQDNPQVNSQLSNHRKEDLOREHS---OKSDSTQVATVADKNNISKSTTN 170  
DB 1316 DASQKATVADNSVANNSESDPKSRRRRSISQDETSADETTAASDETTIADNSKRS 1375  
QY 171 NPNK 174  
DB 1376 KPNR 1379

RESULT 21  
YCGI YEAST STANDARD; PRT; 853 AA.  
AC P25588; P25589; P27513; P87003;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Hypothetical 97.9 kDa protein in CHAI-KRI intergenic region.  
GN YCLO61C OR YCLO61C/YCLO60C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_Taxid=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rasmussen S.W., Von Wettstein D.;  
RL Submitted (MAR-1992) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RA Gromada R.;  
RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 384-596 FROM N.A.  
RC STRAIN=ATCC 28383 / FL100;  
RX MEDLINE=90384830; PubMed=2169608;  
RA Kern L.;  
RT "The URK1 gene of Saccharomyces cerevisiae encoding uridine kinase.";  
RL Nucleic Acids Res. 18:5279-5279(1990).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X59720; CAA42405.1; -;  
DR EMBL; X5398; -; NOT ANNOTATED\_CDS.  
DR PIR; S74279; S74279.  
DR SGD; S0000566; YCLO61C.  
DR GO; GO:0006347; P:chromatin silencing at HML and HMR (sensu S. . .); IGI.  
DR GO; GO:0006348; P:chromatin silencing at telomere; IGI.  
DR GO; GO:0000076; P:DNA replication checkpoint; IGI.  
KW Hypothetical protein.  
FT CONFLICT 505 505  
FT CONFLICT 567 567  
SQ SEQUENCE 853 AA; 97946 MW; 16509FC0BF248D1 CRC64;  
Query Match 9.8%; Score 88; DB 1; Length 853;  
Best Local Similarity 20.6%; Pred. No. 43;  
Matches 40; Conservative 30; Mismatches 80; Indels 44; Gaps 7;

QY 4 VKDPANTTVEKELINKDGEVSELKPRHVTVTIIONGKEMSSIVSEEDFILPVY----- 58  
 Db 256 LEMAEKEKIEVLELLOEI-----LNKKL-----RQKEKPRKEENFOJNANDSGSD 305  
 QY 59 KGELEKGYOPDGEWISGFEFG-----KQDAGVYNLSKDTYIKPVFKKIEEK 104  
 Db 306 SGSESSGSPALSGNEIADYESGSEENDNRRESDEKEDDEILKQKKSHVYKHIINSDSD 365  
 QY 105 KEENKPTDVSKKKNPQVNSQLNESHKEDLQREHSSQKSDSTKQVATVLDKQNI- 163  
 Db 366 TEVEAKP-----KEKADESLEPKRIALNIGHYDNI-----GSDTUKFOETVLDIOTIE 414  
 QY 164 ---SSKSTNNPNK 174  
 Db 415 EWMAERNTIENEVK 428

RESULT 22  
 FENR\_PEA STANDARD; PRT; 360 AA.  
 ID FENR\_PEA  
 AC P10933;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ferredoxin-NADP reductase, leaf isozyme, chloroplast precursor  
 DE (EC 1.18.1.2) (FNR).  
 GN PETH.  
 OS Pisum sativum (Garden pea).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside 1; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.  
 OC NCBI\_TaxID=3888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Little Marvel; TISSUE=Leaf;  
 RA Newman B.J.; Gray J.C.;  
 RT "Characterisation of a full-length cDNA clone for pea ferredoxin-NADP+  
 RT reductase.";  
 RL Plant Mol. Biol. 10:511-520(1988).  
 RN [2]  
 RP SEQUENCE OF 270-360 FROM N.A., AND MUTAGENESIS.  
 RX MEDLINE=93374905; PubMed=8366077;  
 RA Orellano E.G., Calcaterra N.B., Carrillo N., Ceccarelli E.A.;  
 RT "Probing the role of the carboxyl-terminal region of ferredoxin-NADP+  
 RT reductase by site-directed mutagenesis and deletion analysis.";  
 RL J. Biol. Chem. 268:19267-19273(1993).  
 RN [3]  
 RN X-RAY CRYSTALLOGRAPHY (1.70 ANGSTROMS).  
 RX MEDLINE=99396739; PubMed=10467097;  
 RA Deng Z., Aliverti A., Zanetti G., Arakaki A.K., Ottado J.,  
 RA Orellano E.G., Calcaterra N.B., Ceccarelli E.A., Carrillo N.,  
 RA Karpilus P.A.;  
 RT "A productive NADP+ binding mode of ferredoxin-NADP+ reductase  
 RT revealed by protein engineering and crystallographic studies.";  
 RL Nat. Struct. Biol. 6:847-853(1999).  
 CC -1- FUNCTION: MAY PLAY A KEY ROLE IN REGULATING THE RELATIVE AMOUNTS  
 CC OF CYCLIC AND NON-CYCLIC ELECTRON FLOW TO MEET THE DEMANDS OF THE  
 CC PLANT FOR ATP AND REDUCING POWER.  
 CC -1- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized  
 CC ferredoxin + NADPH.  
 CC -1- COFACTOR: FAD.  
 CC -1- PATHWAY: FINAL STEP IN LINEAR PHOTOSYNTHETIC ELECTRON TRANSPORT  
 CC CHAIN: IT HAS ALSO BEEN IMPLICATED IN CYCLIC ELECTRON FLOW AROUND  
 CC PHOTOSYSTEM I, ITS ROLE BEING TO RETURN ELECTRONS FROM FERREDOXIN  
 CC TO THE CYTOCHROME B-F COMPLEX.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST; STROMAL SIDE OF THE THYLAKOID  
 CC MEMBRANE IN THE VICINITY OF THE PHOTOSYSTEM I IN THE NON-STACKED  
 CC AND FRINGE PORTION OF THE MEMBRANE.  
 CC -1- MISCELLANEOUS: FNR IS PROBABLY ATTACHED TO THE MEMBRANE BY A  
 CC SPECIFIC BINDING PROTEIN.  
 CC -1- SIMILARITY: WITH OTHER SPECIES FNR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X12446; CAA30978.1; -;  
 CC EMBL; L15565; AAB59349.1; -;  
 CC EMBL; L15567; AAB59303.1; -;  
 CC EMBL; L15569; AAB59304.1; -;  
 CC PIR; S04030; S04030.  
 CC PDB; 1QFZ; 01-SEP-99.  
 CC PDB; 1QFY; 01-SEP-99.  
 CC PDB; 1QGO; 01-SEP-99.  
 CC PDB; 1QGA; 01-SEP-99.  
 CC InterPro; IPR001709; FPN\_cyt\_reductse.  
 CC InterPro; IPR001433; Oxred\_FAD/NAD(P).  
 CC Pfam; PF00175; NAD\_binding\_1; 1.  
 CC PRINTS; PR00371; FPNCR.  
 CC Oxidoreductase; Flavoprotein; NADP; FAD; Chloroplast; Transit peptide;  
 CC Electron transport; Photosynthesis; Thylakoid; Membrane;  
 CC MultiGene family; 3D-structure.  
 CC TRANSIT 1 52  
 CC CHAIN 53 360  
 CC DISULFID 178 183  
 CC NP BIND 212 230  
 CC MUTAGEN 360 360  
 CC STRAND 68 68  
 CC TURN 71 72  
 CC STRAND 78 78  
 CC TURN 80 81  
 CC STRAND 84 93  
 CC TURN 97 98  
 CC STRAND 103 109  
 CC TURN 111 112  
 CC STRAND 118 119  
 CC TURN 121 125  
 CC STRAND 129 129  
 CC TURN 131 132  
 CC STRAND 135 135  
 CC TURN 139 143  
 CC STRAND 147 148  
 CC TURN 150 151  
 CC STRAND 156 162  
 CC TURN 165 167  
 CC STRAND 169 170  
 CC TURN 173 175  
 CC STRAND 177 184  
 CC TURN 187 188  
 CC STRAND 190 197  
 CC TURN 200 201  
 CC STRAND 204 204  
 CC TURN 207 208  
 CC STRAND 210 216  
 CC TURN 217 220  
 CC STRAND 221 231  
 CC TURN 232 232  
 CC STRAND 243 250  
 CC TURN 253 255  
 CC STRAND 257 258  
 CC TURN 268 268  
 CC STRAND 270 272  
 CC TURN 273 279  
 CC STRAND 280 282  
 CC TURN 284 284  
 CC STRAND 286 287  
 CC TURN 290 290  
 CC STRAND 293 297  
 CC TURN 298 300  
 CC STRAND 301 309  
 CC TURN 311 312



```
FT STRAND 313 319
FT TURN 321 322
FT HELIX 323 337
FT TURN 338 339
FT HELIX 342 351
FT TURN 352 353
FT STRAND 355 360
SQ SEQUENCE 360 AA; 40194 MW; 7F1CC10DEBBA7B24 CRC64;

Query Match 9.8%; Score 87.5; DB 1; Length 360;
Best Local Similarity 24.6%; Pred. No. 19;
Matches 41; Conservative 21; Mismatches 54; Indels 51; Gaps 8;

Qy 4 VKDFARNTT---VKFELINKDTGEV-----SELKP-HRYVTVTIQNGKEM----- 43
Db 148 IGDGDSKTVSLVKRLVYTTDAGEVVKVCSNFLCDLKPGEVKITGPVKGEMLMKPKDP 207
Qy 44 -----STIVSEEDFILPVYKGELEKGYQFDG--WEISGPEGKDKAGYVNLKSDT 92
Db 208 NATVIMLTGTGCIAPFRSFLKMF-FEKHEDYQFNLGLWFLG-----VPTSSL 256
Qy 93 FIKPVFKIEKKKEENKPTDVSKKDNPNVNSHSHRKEDLQ 139
Db 257 LYKEFEKMKKAPENFLDFAVSGE-----QVNDXGKMYIQ 294

RESULT 23
SKM1 YEAST
ID SKM1 YEAST STANDARD; PRT; 655 AA.
AC Q12469; Q06940;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine/threonine-protein kinase SKM1 (EC 2.7.1.-) (Protein kinase
75490 D).
GN SKM1 OR YOL113W OR HRA655.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / GRP88;
RX MEDLINE=97197189; PubMed=9044278;
RA Martin H., Mendoza A., Rodriguez-Pachon J.M., Molina M., Nombela C.;
RT "Characterization of SKM1, a Saccharomyces cerevisiae gene encoding a
novel Ste20/PAK-like protein kinase."
RL Mol. Microbiol. 23:431-444 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96076631; PubMed=7502582;
RA Vandenbol M., Durand P., Portetelle D., Hilger F.;
RT "Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV
including the Ty1-H3 retrotransposon, the sufl(+) frameshift
suppressor gene for trna-Gly, the yeast transfer RNA-Thr-1a and a
delta element."
RL Yeast 11:1069-1075 (1995).
CC -!- FUNCTION: MAY BE INVOLVED IN CELLULAR SIGNALLING OR CYTOSKELETAL
CC FUNCTIONS. MAY PLAY A ROLE IN MORPHOGENETIC CONTROL.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -!- SIMILARITY: Contains 1 CRIB domain.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X69322; CAA49163.1; -.
```

```
DR EMBL; Z48149; CAA88147.1; -.
DR EMBL; Z74855; CAA99132.1; -.
DR PIR; S51884; S51884.
DR SGD; S0005473; SKM1.
DR GO; GO:0000074; P:regulation of cell cycle; IGI.
DR InterPro; IPR000095; PAKbox/RhoBindng.
DR InterPro; IPR001849; PH.
DR InterPro; IPR002290; Ser_Thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00786; PBD; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00285; PBD; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS50108; CRIB; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR Transferase; Serine/threonine-protein kinase; ATP-binding.
KW DOMAIN 3 118
FT DOMAIN 123 136
FT DOMAIN 360 639
FT NP_BIND 366 374
FT BINDING 406 406
FT ACT_SITE 507 507
FT CONFLICT 303 303
FT CONFLICT 306 306
FT CONFLICT 320 320
FT CONFLICT 322 322
FT CONFLICT 452 452
FT CONFLICT 468 468
SQ SEQUENCE 655 AA; 75331 MW; 1F6C8C85FE10D385 CRC64;

Query Match 9.8%; Score 87.5; DB 1; Length 655;
Best Local Similarity 20.5%; Pred. No. 35;
Matches 34; Conservative 36; Mismatches 57; Indels 39; Gaps 7;

Qy 16 FILNKDTGEVSELKPHRTVTVTIQNGKEMSTIVSEEDFILPVYKGE-----EKGQVF 68
Db 209 YIINKRTNSIK---RVSRTLRKKG-----TDSILPVYQSELKPPRPSPDDDKYKF 255
Qy 69 DGEWISGF--EGKDKAGYVNLKSDTPIKPVFKIEKKEE-----ENKPTFDVSK 117
Db 256 TNIEDNKKVREGR-----VHVSKESTADSQTKLGKKEQKVIOSHLRHNDNSTFRPHR 309
Qy 118 KDNPPQV--NHSQLNESHKEDLQREHSHQKSDSDTKDVTATVDKN 161
Db 310 LAPSAPATKHDSKTKWHKEDLLELKNDDSNELIMKMTVAIDVN 355

RESULT 24
GARP PLAFF STANDARD; PRT; 678 AA.
AC P13816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutamic acid-rich protein precursor.
GN GARP.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89040048; PubMed=2903445;
RA Trigilia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.;
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
acid-rich protein (GARP).";
```

RL Mol. Biochem. Parasitol. 31:199-202(1988).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; J03998; AAA29605.1; -

DR PIR; A54514; A54514.

KW Repeat; Malaria; Antigen; Signal.

FT SIGNAL 1 25

FT CHAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.

FT DOMAIN 120 164 15 X 3 AA TANDEM REPEATS OF K-K-X.

FT DOMAIN 372 416 9 X APPROXIMATE TANDEM REPEATS.

FT DOMAIN 417 441 5 X APPROXIMATE TANDEM REPEATS.

FT DOMAIN 576 604 POLY-GLU

FT DOMAIN 605 653 7 X APPROXIMATE TANDEM REPEATS.

FT DOMAIN 654 663 POLY-GLU.

SQ SEQUENCE 678 AA; 2A8F85606496A9E CRC64;

Query Match 9.8%; Score 87.5; DB 1; Length 678;

Best Local Similarity 24.1%; Pred. No. 37; Mismatches 73; Indels 37; Gaps 10;

Matches 46; Conservative 35;

QY 5 KDFARNTTVKFIKND-----TGEV---SELKPHRVTVTIQNGKMSSTIVSEE-----DF 53

DB 25 KCFNGLLNQNLINKSPDSITGRLLNETELEKNK-----DINSKSTLLKKEKDKDD 78

QY 54 ILPVYGELEKGVQPCWISGFEKKGADAGYINLSKDTFKIPVKIKIEKK-----BE 107

DB 79 VPTSDNLKNAHNN--BIS---SSTDPTNIINVNDKDNENSVDKKDKKKGKKKK 133

QY 108 ENKPTDVSCKDNPVNSHNSHREKDLQREHSQ-----KSDSTKDTATVLDKNN 162

DB 134 EKKEKKDKKDKKDKKDKK--EKHKDKKKKENSEVMSLYKGTQGHKPKNATEHGEN 191

QY 163 ISSK--STNN 171

DB 192 LDEMVSEINN 202

RESULT 25

RA50 PYRHO STANDARD; PRT; 879 AA.

AC O5867:

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE DNA double-strand break repair rad50 ATPase.

GN RAD50 OR PH0929.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OX NCBI\_TaxID=53953;

RP SEQUENCE FROM N.A.

RC STRAIN=OT3;

RX MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,

RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,

RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,

RA Funahashi T., Tanaka T., Kudo H. Y., Yamazaki J., Kishida N., Oguchi A.,

RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,

RA Masuchi Y., Shizuya H., Kikuchi H.;

RT "Complete sequence and gene organization of the genome of a hyper-

RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.;"

RL DNA Res. 5:55-76(1998).

CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The

CC rad50/mre11 complex possesses single-strand endonuclease activity

CC and ATP-dependent double-strand-specific exonuclease activity.

CC Rad50 provides an ATP-dependent control of mre11 by unwinding

CC and/or repositioning DNA ends into the mre11 active site (By

CC similarity).

CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).

CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AP000004; BAA30025.1; -

DR PIR; C71083; C71083.

DR HAMAP; MF\_00449; -; 1.

DR InterPro; IPR003593; AAA ATPase.

DR InterPro; IPR003439; ABC transporter.

DR InterPro; IPR003405; SMC\_C.

DR InterPro; IPR003395; SMC\_N.

DR Pfam; PF04423; Rad50\_zn\_hook; 1.

DR Pfam; PF02483; SMC\_C; 1.

DR Pfam; PF02463; SMC\_N; 1.

DR SMART; SM00382; AAA; 1.

KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.

FT NP\_BIND 30 37 AIP (BY SIMILARITY).

FT DOMAIN 141 744 COILED COIL (POTENTIAL).

SQ SEQUENCE 879 AA; 103673 MW; FD4E30FD1BBCB29 CRC64;

Query Match 9.8%; Score 87.5; DB 1; Length 879;

Best Local Similarity 21.0%; Pred. No. 49; Mismatches 51; Indels 27; Gaps 3;

Matches 30; Conservative 35;

QY 4 VKDFARNTTVKFIKNDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELE 63

DB 551 LNDYKNESTKLEIETDKAKKELSEIEDRLRLGPKTIDELSGRI-----RELE 598

QY 64 KYQFDGWEISGFEKKGADAGYINLSKDTFKIPVKIKIEKKENKTFDVKKKDKNPQ 123

DB 599 K-----PHNK---YIEAKNAEKELDILSKDERELDKAPEELAKIETDIE 643

QY 124 VNSHNSHREKDLQREHSQ 146

DB 644 KVTSQLQKFKDQKYEKKRE 666

RESULT 26

UN89 CAEEL STANDARD; PRT; 6632 AA.

AC O01761; Q17362;

DT 15-SEP-2003 (Rel. 42, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).

GN UNC-89 OR C09D1.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.

RC STRAIN=Bristol N2;

RX MEDLINE=96180278; PubMed=8603916;

RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;

RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line

RT assembly, encodes a giant modular protein composed of Ig and signal

RT transduction domains.;"

RL J. Cell Biol. 132:835-848(1996).

RP [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Du Z., Le T.T., Wilson R.;

```

RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBI-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U33058; AAB00542.1; -;
CC EMBL; AF003131; AAB54132.2; -;
CC PDB; 1PHO; 20-DEC-00.
CC WormPep; C09D1.1; CE30426.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR007110; IG-like.
CC InterPro; IPR003598; IG_c2.
CC InterPro; IPR003306; IG_MHC.
CC InterPro; IPR001849; PH.
CC InterPro; IPR007850; RCSD.
CC InterPro; IPR000219; RhGEF.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00041; fn3; 1.
CC Pfam; PF00047; ig; 47.
CC Pfam; PF00165; PH; 1.
CC Pfam; PF05177; RCSD; 5.
CC Pfam; PF00621; RhGEF; 1.
CC Pfam; PF00018; SH3; 1.
CC SMART; SM00408; IGC2; 23.
CC SMART; SM00325; RhGEF; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS50010; DH 2; 1.
CC PROSITE; PS50835; IG LIKE; 49.
CC PROSITE; PS50003; PH DOMAIN; 1.
CC PROSITE; PS50002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.

```

Query Match 9.7%; Score 87; DB 1; Length 6632;  
 Best Local Similarity 21.8%; Pred. No. 4.7e+02;

```
Matches 37; Conservative 33; Mismatches 80; Indels 20; Gaps 7;
QY 11 TTWKFIKMGGEVSELSKPHRVTVTIQNGKMSSTIVSEEDF--ILPVYKGELEKGVQF 68
Db 1127 TRAKLTQVKKPPAFEDLRP--VSLTVKGS---AVFSAHAFGIPLPYENSVNGRKVR 1181
QY 69 DQWE---ISGFGKDGAGVIVNLSKDTFTKPV-----FKKIEKK--EENKFTFDVSKKK 119
Db 1182 DQEGVATRDSTVDGASITDITATYYSEVNHITISVAENTLGAETGAQLTIEPKK 1241
QY 120 DNPQVNHLSQNSHREKEDLQREHSQKSDSTKDTATVLDKNNISKSTT 169
Db 1242 ESVVVEKQLSSS-----EVQKIAQOVKEASPEATTITMETSLSITKTT 1287

RESULT 27
ID RAT1 YEAST STANDARD; PRT; 1006 AA.
AC Q02752;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ribonucleic acid trafficking protein 1 (5'-3' exoribonuclease)
DE (EC 3.1.11.-) (P116)
GN RAT1 OR HK1 OR TAP1 OR YOR048C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92311925; PubMed=1628825;
RA Amberg D.C., Goldstein A.L., Cole C.N.;
RT "Isolation and characterization of RAT1: an essential gene of
RT Saccharomyces cerevisiae required for the efficient nucleocytoplasmic
RT trafficking of mRNA."
RL Genes Dev. 6:1173-1189(1992).
RN [2]
RP SEQUENCE FROM N.A., AND EXORIBONUCLEASE ACTIVITY.
RX MEDLINE=93109318; PubMed=8417335;
RA Kenna M., Stevens A., McCammon M., Douglas M.G.;
RT "An essential yeast gene with homology to the exonuclease-encoding
RT XRN1/KEM1 gene also encodes a protein with exoribonuclease
RT activity."
RL Mol. Cell. Biol. 13:341-350(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93268292; PubMed=8497260;
RA Aldrich T.L., di Segni G., McConaughy B.L., Keen N.J., Whelen S.,
RA Hall B.D.;
RT "Structure of the yeast TAP1 protein: dependence of transcription
RT activation on the DNA context of the target gene."
RL Mol. Cell. Biol. 13:3434-3444(1993).
RN [4]
RP SEQUENCE FROM N.A.
RL Landt O., Hiesel R., Unseld M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Bohn C., Bolotin-Fukuhara M., Daignan-Fornier B., Dang D.V.,
RA Valens M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May function in the processing and/or trafficking of
CC nuclear mRNA. May be involved in general transcription as well.
CC Possesses 5'-3' exoribonuclease activity degrading poly(A) to
CC mainly 5'-AMP.
CC -!- COFACTOR: Requires magnesium.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the 5'-3' exonuclease family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entitles requires a license agreement. Usage by and for commercial
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; S61567; AAB26818.1; -
CC EMBL; M95626; AAA34960.1; -
CC EMBL; L06011; AAA16950.1; -
CC EMBL; Z11746; -; NOT ANNOTATED CDS.
CC EMBL; Z74956; CAA99240.1; -
CC PIR; S20126; S20126.
CC SGD; S0005574; RAT1.
CC GO; GO:0005634; C:nucleus; IDA
CC GO; GO:0004534; F:5'-3' exoribonuclease activity; IDA.
CC GO; GO:0006365; P:35S primary transcript processing; IMP.
CC GO; GO:0006396; P:RNA processing; IMP.
CC InterPro; IPR004859; Put 53exo.
CC Pfam; PF03159; XRN N; 1.
CC Nuclear protein; Hydrolase; Nuclease; Exonuclease; Repeat.
CC DOMAIN 525 528 POLY-GLU.
CC DOMAIN 793 797 POLY-ASN.
CC DOMAIN 955 999 CONTAINS 2 X SRYD, 2 X NNNY, AND 2 X
CC FT MUTAGEN 683 YSGN REPEATS.
CC FT Y->H: IN ALLELE TAP1-1; ACTIVATES
CC FT TRANSCRIPTION OF THE PROMOTER-DEFECTIVE
CC FT YEAST SUP4 TRNA(TYR) ALLELE SUP4A53T61.
SQ SEQUENCE 1006 AA; 115933 MW; 5DD5B0245F3E12A CRC64;
Query Match 9.6%; Score 86.5; DB 1; Length 1006;
Best Local Similarity 19.0%; Pred. No. 66;
Matches 32; Conservative 40; Mismatches 47; Indels 49; Gaps 8;
QY 37 IQNGKMSSTIVSEEDFILPVY--KGELEKGVQFQWEISGFEKKDAGYVNLKDTFI 94
Db 415 MSKGQDEHPTVATEQ---LQMYDTQCNLAGK---SMNLTTSD-----MVRLKKELML 460
QY 95 -----KPVFKIEEKKKEENKPTEDVSKKKNPQNVHSQ-----LNESHK 135
Db 461 ANEGNEEAIKVKQSDKNNELMKDISKEEDDVAISKANKTNFNAEVMKQIKKHL 520
QY 136 EDLQREEH-----SOKSDSTKDTATVLDK-----NNISSKSTT 169
Db 521 EKDNEEBIIAKDSKKVKTKEASECDLDAEIKDEIVADVNDRENSSETT 568
RESULT 28
RBP2 PLAVB STANDARD; PRT; 1251 AA.
AC Q00789;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 2 (Fragment).
GN RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=923115338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merozoites."
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M88098; AAA29744.1; -  
 KW Malaria; Receptor; Membrane.  
 FT NON\_TER 1251 1251  
 SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 9.6%; Score 86.5; DB 1; Length 1251;  
 Best Local Similarity 22.8%; Pred. No. 84;  
 Matches 46; Conservative 40; Mismatches 77; Indels 39; Gaps 10;

Qy 2 IWKDFARNTTVKPF-----ILNKD---TGEVSELKPH--RVTVTTIQNGKMSSTI--- 47  
 Db 234 IALEDAQIDTEVSKIEQINREIMNKDKBIKSYLSEIKYKDKCTTEISNSKRGDKIEFL 293  
 Qy 48 -----VSEEDFLPVYKGELEKGYQPDGWEISGPE-GKKDAGYVNL--SKDTFKPVFKK 100  
 Db 294 EKFPNEESNSKVNINEINIRNSEQYLKDIEDAEKQASTKVELFKHKTETISNIFKE 353  
 Qy 101 -----IBEKKEENKPTPDVSK--KKONPOV-----NHSQNSHSHKEDLQREHS 144  
 Db 354 SEILGVETYSOKKINKAEDIMKEIERHNSHSEIQTQVKGFQENLNKLNPHNYDNADELNN 413  
 Qy 145 QKSDSTKDVATVLD--KNNIS 164  
 Db 414 DKSTNAKVLITNLESVKHNL 435

# RESULT 29

SYST LYES  
 ID SYST LYES STANDARD; PRT; 200 AA.  
 AC P27058;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Systemin precursor.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Leaf.  
 RX MEDLINE=92196587; PubMed=1549783;  
 RA McGurl B., Pearce G., Orozco-Cardenas M., Ryan C.A.;  
 RT "Structure, expression, and antisense inhibition of the systemin  
 precursor gene.";  
 RL Science 255:1570-1573 (1992).  
 RN [2]  
 RP SEQUENCE OF 179-196.  
 RA Pearce G., Strydom D., Johnson S., Ryan C.A.;  
 RT "A polypeptide from tomato leaves induces wound-inducible proteinase  
 inhibitor proteins.";  
 RL Science 253:895-898 (1991).  
 RN [3]  
 RP FUNCTION.  
 RX MEDLINE=20175266; PubMed=10708853;  
 RA Ryan C.A.;  
 RT "The systemin signaling pathway: differential activation of plant  
 defensive genes";  
 RL Biochim. Biophys. Acta 1477:112-121 (2000).  
 CC -!- FUNCTION: ACTIVATES A LIPID-BASED SIGNAL TRANSDUCTION PATHWAY IN  
 CC WHICH LINOLENIC ACID IS CONVERTED TO JASMONIC ACID, A POTENT  
 CC ACTIVATOR OF DEFENSE GENE TRANSCRIPTION, INCLUDING PROTEINASE  
 CC INHIBITOR.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- TISSUE SPECIFICITY: ALL ORGANS EXCEPT THE ROOTS. TRANSPORTED OUT  
 CC OF WOUNDS TO DISTAL TISSUES.  
 CC -!- INDUCTION: By wounding; in leaves.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M84800; AAA34182.1; -  
 DR EMBL; M84801; AAA34184.1; -  
 DR PIR; T07149; T07149.  
 KW Hormone; Repeat.

FT PROPEP 1 178  
 FT CHAIN 179 196  
 FT PROPEP 197 200  
 FT REPEAT 3 8  
 FT REPEAT 37 45  
 FT REPEAT 80 88  
 FT REPEAT 117 125  
 FT REPEAT 145 153  
 SQ SEQUENCE 200 AA; 22999 MW; FA251D94BAA9C5A9 CRC64;

Query Match 9.6%; Score 86; DB 1; Length 200;  
 Best Local Similarity 21.5%; Pred. No. 13;  
 Matches 37; Conservative 30; Mismatches 59; Indels 46; Gaps 8;

Qy 1 KIVVK-----DFARNTTVKGFILNKOTGEVSELKPHRVTVTTIQNGKMSSTIVSEDFIL 55  
 Db 34 KIEKETPSQDINNKTIISSYVLDDTQEIPEH-H-----EGGYVKEKIVEKETISQ 86  
 Qy 56 PVYKGELEKGYQPDGWEISGFEKGDAGYVNLISKDTFKIPVFKKIEKKE---EENKPT 112  
 Db 87 YIHK-----IEGDDA-----QELKVEYEVEEVEKEKIVEKETPS 122  
 Qy 113 FDVSKKKDN-----POVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDK 160  
 Db 123 QDINNKGDDAQEKPKVEH---REGDDKTPSQDIIRMEGEGALEITKVVCEK 171

# RESULT 30

GLNA\_PIRKO  
 ID GLNA\_PIRKO STANDARD; PRT; 443 AA.  
 AC O08467;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
 GN GLNA.  
 OS Pyrococcus kodakaraensis.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Thermococcus.  
 OX NCBI\_TaxID=69014;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KOD1;  
 RX MEDLINE=97316461; PubMed=9172372;  
 RA Adul Rahman R.N.Z., Jongsareejit B., Fujiwara S., Imanaka T.;  
 RT "Characterization of recombinant glutamine synthetase from the  
 RT hyperthermophilic archaeon Pyrococcus sp. strain KOD1.";  
 RL Appl. Environ. Microbiol. 63:2472-2476 (1997).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
 CC L-glutamine.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D86222; BAA20530.1; --  
 DR HSP; P06201; ILGR.  
 DR InterPro; IPR001691; GLN synth.  
 DR InterPro; IPR004809; GLN.  
 DR InterPro; IPR001637; GLN adenyltn.  
 DR Pfam; PF00120; gln-synt\_1.  
 DR Pfam; PF03951; gln-synt\_N; 1.  
 DR ProDom; PD001057; Gln synt\_C; 1.  
 DR TIGRFAMs; TIGR00653; GLN; 1.  
 DR PROSITE; PS00180; GLNA 1; 1.  
 DR PROSITE; PS00181; GLNA\_ATP; 1.  
 KW Ligase.  
 FT BINDING 362 362 AMP (UNDER CONDITIONS OF ABUNDANT  
 FT GLUTAMINE) (BY SIMILARITY).  
 SQ SEQUENCE 443 AA; 50259 MW; 9426DCCFEF18168 CRC64;  
 Query Match 9.6%; Score 86; DB 1; Length 443;  
 Best Local Similarity 17.8%; Pred. No. 30;  
 Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;  
 QY 58 YKELEKGYQDFGWEISGPEKGDAGYVINSKDTFI----- 94  
 DB 42 YEAEVDGVSFDGSSIPGEGIEDSLIFKADPSTVAEIPWEGIGRVGYIVKGDEPYQA 101  
 QY 95 --KPVFKTEERKEEN-----KPTFDVSKKD----- 120  
 DB 102 DPGILKRVLRLEKGLKAHIGPEPEFYIFKNGTWELHDPDGGYFDVLGLDKAREIR 161  
 QY 121 -----NPOVNSQLNESHRKEDLQREHSOKSD---STKDVATATVLD 159  
 DB 162 REIALYMPYLGLKPEVLHHEVGAQHEIDFRIDEALRTADNIVSKHVKAQAE 215  
 RESULT 31  
 PBPA\_BACSU STANDARD; PRT; 914 AA.  
 ID PBPA\_BACSU STANDARD; PRT; 914 AA.  
 AC P39793;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Penicillin-binding protein 1A/1B (PBPI) [Includes: Penicillin-  
 DE insensitive transglycosylase (EC 2.4.2.-) (Peptidoglycan TGase);  
 DE Penicillin-sensitive transpeptidase (EC 3.4.-.-) (DD-transpeptidase)].  
 GN PONA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]\_TaxID=1423;  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 499-515.  
 RC STRAIN=168;  
 RX MEDLINE=95113769; PubMed=7814321;  
 RA Popham D.L., Setlow P.;  
 RT "Cloning, nucleotide sequence, and mutagenesis of the Bacillus  
 RT subtilis *ponA* operon, which codes for penicillin-binding protein  
 RT (PBP) 1 and a PBP-related factor.";  
 RL J. Bacteriol. 177:326-335 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / Marburg;  
 RX MEDLINE=96349105; PubMed=8760912;  
 RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,  
 RA Serror P.;  
 RT "Sequence analysis of the Bacillus subtilis chromosome region between  
 RT the *serA* and *gld* loci cloned in a yeast artificial chromosome.";  
 RL Microbiology 142:2005-2016 (1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portecelle D., Forwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche M., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis.";  
 RN Nature 390:249-256 (1997).  
 RN [4]  
 RP GROWTH REQUIREMENTS.  
 RC STRAIN=168;  
 RX MEDLINE=98389671; PubMed=9721295;  
 RA Murray T., Popham D.L., Setlow P.;  
 RT "Bacillus subtilis cells lacking penicillin-binding protein 1 require  
 RT increased levels of divalent cations for growth.";  
 RL J. Bacteriol. 180:4555-4563 (1998).  
 RN [5]  
 RP SUBCELLULAR LOCATION.  
 RC STRAIN=168;  
 RX MEDLINE=99255546; PubMed=10322023;  
 RA Pedersen L.B., Angert E.R., Setlow P.;  
 RT "Septal localization of penicillin-binding protein 1 in Bacillus  
 RT subtilis.";  
 RL J. Bacteriol. 181:3201-3211 (1999).  
 CC -!- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED  
 CC PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A  
 CC PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN  
 CC (FORMATION OF LINEAR GLYCAN STRANDS) AND A PENICILLIN-SENSITIVE  
 CC TRANSEPTIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE  
 CC SUBUNITS) (BY SIMILARITY).  
 CC -!- PATHWAY: Peptidoglycan synthesis; final stages.  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, PROBABLY FOUND ALL  
 CC OVER THE WHOLE CELL AT LOW CONCENTRATIONS. ALSO LOCALIZES TO THE  
 CC DIVISION SITE IN VEGETATIVE CELLS.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSION IS CONSTANT DURING GROWTH,  
 CC DECREASES DURING SPOULATION AND IS INDUCED APPROXIMATELY 15 MIN  
 CC INTO SPORE GERMINATION.  
 CC -!- PTM: THE PRODUCT EXPRESSED FROM THE TRANSLATION OF THE *PONA* GENE  
 CC APPEARS AS TWO BANDS ON A GEL (1A AND 1B), BUT THE SPECIFIC AMINO  
 CC ACID SEQUENCE OF EACH PROTEIN IS UNKNOWN.  
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -!- MISCELLANEOUS: CELLS LACKING THE PROTEIN REQUIRE INCREASED LEVELS  
 CC OF MG(2+) OR CA(2+) FOR GROWTH AND GERMINATION. APPROXIMATELY 50%  
 CC OF CELLS WITHOUT THE PROTEIN CONTAIN ABNORMAL FT2Z RINGS,  
 CC SUGGESTING IT IS INVOLVED IN SEPTUM SYNTHESIS; INCREASED LEVELS OF  
 CC MG(2+) OR CA(2+) ONLY PARTIALLY ELIMINATE THE SEPTATION DEFECTS.  
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE  
 CC TRANSGLYCOSYLASE FAMILY.  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE  
 CC TRANSEPTIDASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



```

QY 142 EHSQSDS 149
DB 4747 EDEKSDS 4754

RESULT 33
DRS1_YEAST
ID DRS1_YEAST STANDARD; PRT; 752 AA.
AC P32892;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable ATP-dependent RNA helicase DRS1.
GN DRS1 OR YLI008W OR L1345.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93087480; PubMed=1454790;
RA Ripmaster T.L., Vaughn G.P., Woolford J.L. Jr.;
RT "A putative ATP-dependent RNA helicase involved in Saccharomycetes
RT cerevisiae ribosome assembly."
RL Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / FY23;
RX MEDLINE=96405918; PubMed=8810043;
RA Miosga T., Zimmermann F.K.;
RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT a 43.7 kb fragment of chromosome XII including an open reading frame
RT homologous to the human cystic fibrosis transmembrane conductance
RT regulator protein CFTR."
RL Yeast 12:693-708 (1996).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansoerge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Newtich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Reckmann S., Rieger M., Rinke M., Rose M.,
RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz P.,
RA Underwood A.P., Ureterazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hohelsel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII."
RL Nature 387:87-90 (1997).
CC -!- FUNCTION: PROBABLE HELICASE INVOLVED IN RIBOSOME ASSEMBLY.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L00683; AAA34666.1; -
DR EMBL; X21488; CAA62783.1; -
DR EMBL; Z73113; CAA97452.1; -
DR FIR; S64750; S64750.
DR HSP; Q58083; 1HV8.
DR SGD; S0003931; DRS1.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000629; DEAD_box.
DR InterPro; IPR001650; Helicase_C.

```

Pfam: PF00270; DEAD; 1.  
Pfam: PF00271; helicase\_C; 1.  
SMART: SM00487; DEXDC; 1.  
SMART: SM00490; HELICC; 1.  
KW PROSITE; PS00039; DEAD ATP HELICASE; 1.  
KW ATP-binding; RNA-binding; Helicase; Nuclear protein.  
FT DOMAIN 170 190 POLY-GLU.  
FT NP\_BIND 275 282 ATP (POTENTIAL).  
FT SITE 385 388 DEAD BOX.  
FT CONFLICT 1 68  
SQ SEQUENCE 752 AA; 84843 MW; 60747607A6E5E48 CRC64;  
MVVGTKYKYNLDFVFTISDSDDVPILSDSDDEKVEAKKTT  
KKRGKNNKKVSEGNLDVHEDLD -> MTKKSLRLRL  
RRSGVRITRKLRLVIGITSMRMFMRTWT (IN REF. 1).  
Query Match 9.5%; Score 85.5; DB 1; Length 752;  
Best Local Similarity 21.8%; Pred. No. 57;  
Matches 38; Conservative 29; Mismatches 64; Indels 43; Gaps 8;  
QY 17 ILNKDTEVSELKPHRVTVTIQNGKMSSTTVSEEDFILPVYKGELEKGYQFD----- 69  
DB 26 ILDSDDDEKVEAKK---TKKRGKN-NKKVSEGNLDVHEDLDAGKFDLADDTT 81  
QY 70 ---GWEI--SGFEKGDAGVYVNLKDTFKIPVFK-----IBEKKEENKPTFD 114  
DB 82 SNFGQWNLAEAGESNKKDAEAFV--KKDVLDKIIRKGLVGMHIDSKQEETEKE-K 138  
QY 115 VSKKDN-----PQVNHSQLNSHRKEDLQREHHSOKSDSTKDTV 154  
DB 139 VEKENDSDDEALMDGFGMGAPMNGDENQSEEEEEKEEEEEEEBEQEMT 192  
RESULT 34  
RGAL\_YEAST  
ID RGAL\_YEAST STANDARD; PRT; 1007 AA.  
AC P39083; P39934;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Rho-type GTPase-activating protein 1.  
GN RGAL OR DBM1 OR THE1 OR YOR127W OR O3290 OR YOR3290W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
RX STRAIN=S288c;  
RX MEDLINE=96239492; PubMed=8657111;  
RA Chen G.-C., Zheng L., Chan C.S.M.;  
RT "The LIM domain-containing Dbm1 GTPase-activating protein is required  
RT for normal cellular morphogenesis in Saccharomyces cerevisiae.";  
RL Mol. Cell. Biol. 16:1376-1390 (1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=S288c / FY1679;  
RX MEDLINE=97060020; PubMed=8904341;  
RA Wiemann S., Reckmann S., Benes V., Voss H., Schwager C., Vlcek C.,  
RA Stegemann J., Zimmermann J., Exfle H., Paces V., Ansoerge W.;  
RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV  
RT from Saccharomyces cerevisiae reveals 30 open reading frames.";  
RL Yeast 12:281-288 (1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97344368; PubMed=9200815;  
RA Voss H., Benes V., Andrade M.A., Valencia A., Reckmann S., Teodoru C.,  
RA Schwager C., Paces V., Sander C., Ansoerge W.;  
RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV.";  
RL Yeast 13:655-672 (1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX STRAIN=S288c;  
RX MEDLINE=96101594; PubMed=7498791;  
RA Stevenson B.J., Ferguson B., de Virgilio C., Bi E., Pringle J.R.,





```

Query Match          9.5%; Score 85.5; DB 1; Length 2459;
Best Local Similarity 24.4%; Pred. No. 2e+02;
Matches 40; Conservative 26; Mismatches 53; Indels 45; Gaps 8;

14 KEFILNDKTGVSELKPHRVTVTTQNGKEMSTIVSEEDFILPVYKGELEKYGQFDGWEI 73
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
583 EKVIWKDKPGKVESKPSVTEKEVPSKEEQS-----PV-KAEVA-----EK 622

74 SGFEGKKDAGVYINLSKDTFTKPVFK-KLEBKKEENKPTFDVSKKKDNDQVNHSQLNES 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
623 AATESKP-----KVTKDKVVYKGEIKTKPEEKGE--KPKKEVAKED----- 662

133 HRKEDLQREHSQKSDSTKYDTATVL--DKNNISKSTTNNPNK 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
663 --KTPLEKDEKPKKEAAKKEIKKEIKKEEKELKKEVKKETPLK 704

```

[illegible]

RESULT 38

```
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C106.14c in chromosome II.
GN SPBC106.14C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Spouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Coffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880 (2002).
CC -!- SIMILARITY: TO YEAST YGR245C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL10295; CAB53730.1; -.
DR F1R; T39271; T39271.
DR GenBank; SPBC106.14c; -.
DR Pfam; PF05285; SDAL; 1.
KW Hypothetical protein.
FT DOMAIN 250 254 POLY-GLU.
FT DOMAIN 559 562 POLY-GLU.
SQ SEQUENCE 719 AA; 81250 MW; 93864D55B86BD1D6 CRC64;
Query Match 9.5%; Score 85; DB 1; Length 719;
Best Local Similarity 23.9%; Pred. No. 59;
Matches 34; Conservative 32; Mismatches 62; Indels 14; Gaps 5;
QY 15 EFLNKDGTGEVSELKPHRVTVITQNGKMSSTIVSEDFILPVYKGEI-----EKGYQFDG 70
Db 469 DMLKRRDGRKGLASIEKMDRT-PLKYGEELNVTHGIQLELLAQYKAEGEEGEGDDWDN 527
QY 71 WEIS-GFEKKDAGYVINLSKDTIFKPVFKTEEEKKEENKPTFDVSKSKDNPQVHSQL 129
Db 528 WEVSEDDQNSDDSGGWIDVSDNNI-----ELSDSDEEEKAT---ARKESDEKSSQK 579
QY 130 NESHKREDLQREHSQKSDSTK 151
DB 580 ELVDRMTELASQSILTPNDLKK 601
RESULT 40
ID IF2_STAEP STANDARD; PRT; 720 AA.
AC Q8CST4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Translation initiation factor IF-2.
DE Translation initiation factor IF-2.
GN INFB OR SE0945.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016747; AA004542.1; -.
DR HAMAP; MF_00100; -.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR Pfam; PF04760; IF2_N; 2.
DR ProDom; PD186100; IF2; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 225 373 G-DOMAIN.
FT NP_BIND 231 238 GTP (BY SIMILARITY).
FT NP_BIND 277 281 GTP (BY SIMILARITY).
FT NP_BIND 331 334 GTP (BY SIMILARITY).
SQ SEQUENCE 720 AA; 79343 MW; 07FBSA6A59CF970C CRC64;
Query Match 9.5%; Score 85; DB 1; Length 720;
Best Local Similarity 28.0%; Pred. No. 59;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;
QY 101 IEEKKEENKPTFDVSKKD---NPQVNHSQLNESHKEDLQREHSQKSDSTK---D 152
Db 38 LEEEQIKALDKKFKASQAKDTNKQNTQNNHQSNNKQNSNDKQKQSKNNSKPTKKQGN 97
QY 153 VTATVLDKNNISSKSTTNPNK 174
Db 98 NKGKQNNKNNKNNKNNKNNK 119
Search completed: February 10, 2004, 10:49:56
Job time : 9.08902 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 35.0346 Seconds  
(without alignments)  
697.707 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799

Sequence: 1 DTGVSBLKPHRVTVTQNG.....ATVLDKNNISKSTNNPNK 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

- 1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	799	100.0	773	22 AAB48343	S. pneumoniae Spi3
2	799	100.0	2120	21 AAY81710	Streptococcus pneu
3	799	100.0	2140	24 ABU01020	S. pneumoniae type
4	615	77.0	117	19 AAW55096	Streptococcus pneu
5	615	77.0	117	23 ABP54590	S. pneumoniae Sp04
6	118	14.8	746	22 AAG81779	S. epidermidis ope
7	118	14.8	778	23 ABP39023	Staphylococcus epi
8	107.5	13.5	484	21 AAG47777	Arabidopsis thalia
9	105	13.1	665	21 AAB18278	Plasmodium falci

10	101.5	12.7	564	22 ABB61977	Drosophila melanog
11	96.5	12.1	2060	23 AAE20967	Staphylococcus lug
12	96	12.0	2500	21 AAB18272	Plasmodium falci
13	93.5	11.7	408	21 AAG37134	Arabidopsis thalia
14	93.5	11.7	456	21 AAG37133	Arabidopsis thalia
15	93.5	11.7	476	21 AAG37132	Arabidopsis thalia
16	92.5	11.6	2519	22 ABG16636	Novel human diagn
17	90.5	11.3	209	21 AAB42897	Human ORFX ORF2661
18	90.5	11.3	757	21 AAY44364	Human cell cycle r
19	90.5	11.3	758	21 AAB53319	Human colon cancer
20	90	11.3	442	23 ABP38188	Staphylococcus epi
21	90	11.3	645	24 ABP38188	Pathogen specific
22	90	11.3	654	24 ABP56879	Staphylococcus epi
23	90	11.3	1183	22 ABB58769	Drosophila melanog
24	89.5	11.2	402	20 AAY03190	S. aureus trigger
25	89.5	11.2	511	20 AAY35091	Chlamydia pneumoni
26	89.5	11.2	525	20 AAY03189	S. aureus trigger
27	89.5	11.2	1408	22 ABB58704	Drosophila melanog
28	89	11.1	645	24 ABP18979	Pathogen specific
29	88.5	11.1	281	22 ABB64828	Drosophila melanog
30	88.5	11.1	1141	22 AAG85008	Shrimp white spot
31	88	11.0	225	23 ABP73992	Candida albicans e
32	88	11.0	258	22 AAB94584	Human protein sequ
33	88	11.0	817	22 AAM79318	Human protein SEQ
34	88	11.0	817	22 AAM79318	Human protein SEQ
35	87.5	11.0	635	23 ABP73209	Candida albicans e
36	87.5	11.0	2515	22 ABB12281	Human secreted pro
37	87.5	11.0	2515	22 AAM80268	Human protein SEQ
38	87.5	11.0	2515	22 AAM80269	Human protein SEQ
39	87.5	11.0	3021	24 ABP75873	Human secretory po
40	87	10.9	313	21 AAG47057	Arabidopsis thalia
41	87	10.9	607	22 AAM39950	Human polypeptide
42	87	10.9	635	21 AAB18176	Plasmodium falci
43	87	10.9	902	23 ABP40312	Staphylococcus epi
44	87	10.9	922	22 AAG85023	Shrimp white spot
45	86.5	10.8	443	22 AAU38439	Salmonella typhi c

ALIGNMENTS

RESULT 1

AAB48343

ID AAB48343 standard; Protein; 773 AA.

XX AAB48343;

AC AAB48343;

XX 20-APR-2001 (first entry)

DT S. pneumoniae Spi30 polypeptide.

DE Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;

KW bronchial; lung; blood; infection; immune response; immunotherapy;

KW antibacterial; auditory; vaccine.

XX Streptococcus pneumoniae.

OS WO200076540-A2.

XX 21-DEC-2000.

PF 09-JUN-2000; 2000WO-US15925.

PR 10-JUN-1999; 99US-0138453.

XX (MEDI-) MED IMMUNE INC.

 XX Adamou JE, Choi GH; |

PT preventing pneumococcal infections, particularly infections caused by  
 PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
 XX blood infections  
 PS Claim 8; Page 51-54; 54pp; English.  
 XX  
 CC The invention relates to novel immunogenic polypeptides, Sp128 and Sp130  
 CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
 CC the treatment and prevention of pneumococcal infections, particularly  
 CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
 CC bronchial, lung or blood infections. The antigens are used as immunogenic  
 CC agents to stimulate an immune response. The antisera and antibodies may  
 CC also be used in diagnosing and treating pneumococcal infections.  
 CC Recombinant polypeptides serve as a mechanism for stimulating production  
 CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
 CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the S. pneumoniae Sp130 polypeptide.  
 XX  
 XX Sequence 773 AA;

Query Match 100.0%; Score 799; DB 22; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-71;  
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 60  
 DB 620 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 679  
 QY 61 DAGYVNLKSDTKDVTATVLDKNNISSKSTNNPNK 154  
 DB 680 DAGYVNLKSDTKDVTATVLDKNNISSKSTNNPNK 739  
 QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
 DB 740 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 773

RESULT 2  
 AAY81710  
 ID AAY81710 standard; Protein; 2120 AA.  
 XX  
 AC AAY81710;  
 XX  
 DT 02-JUN-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae protein sequence ID3.  
 XX  
 KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200006738-A2.  
 XX  
 PD 10-FEB-2000.  
 XX  
 PF 27-JUL-1999; 99WO-GB02452.  
 XX  
 PR 27-JUL-1998; 98GB-0016336.  
 PR 19-MAR-1999; 99US-0125329.  
 XX  
 PA (MICR-) MICROBIAL TECHNIQS LTD.  
 XX  
 PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
 XX  
 DR WPI; 2000-195301/17.  
 DR N-PSDB; AA291806.  
 XX  
 PT Streptococcal proteins and polynucleotides useful for diagnosis,  
 PT treatment and prophylaxis of bacterial infections  
 XX

PS Claim 2; Page 41-42; 76pp; English.  
 XX  
 CC This sequence represents a Streptococcus pneumoniae protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful as vaccines and also in  
 CC diagnostic assays. The sequences are useful for the detection or  
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
 CC infection. As the sequences can be used to treat S. pneumoniae infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
 CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
 CC meningitis.  
 XX  
 XX Sequence 2120 AA;

Query Match 100.0%; Score 799; DB 21; Length 2120;  
 Best Local Similarity 100.0%; Pred. No. 9.3e-71;  
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 60  
 DB 1933 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 1992  
 QY 61 DAGYVNLKSDTKDVTATVLDKNNISSKSTNNPNK 154  
 DB 1993 DAGYVNLKSDTKDVTATVLDKNNISSKSTNNPNK 2086  
 QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
 DB 2053 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2086

RESULT 3  
 ABU01020  
 ID ABU01020 standard; Protein; 2140 AA.  
 XX  
 AC ABU01020;  
 XX  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain protein from coding region #590.  
 XX  
 KW Bacterial meningitis; pneumonia; sepsis; otitis media;  
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KW auditory; respiratory; gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae type 4 strain.  
 XX  
 PN WO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-IB02163.  
 XX  
 PR 27-MAR-2001; 2001GB-0007658.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Masignani V, Tettelin H, Fraser C;  
 XX  
 DR WPI; 2003-040579/03.  
 DR N-PSDB; ABX06302.  
 XX  
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
 PT



PA (BARA/) BARASH S C.  
 PA (DILL/) DILLON P J.  
 PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX  
 DR WPI; 2002-479261/51.  
 DR N-PSDB; ABQ84825.  
 XX  
 DR New Streptococcus pneumoniae antigens, useful for detecting  
 PT Streptococcus and for preventing or attenuating disease caused by  
 PT Streptococcus infection -  
 XX  
 PS Claim 11; Page 29; 70pp; English.  
 XX  
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
 CC The S. pneumoniae antigens have antibacterial activity and can be  
 CC used in vaccines. The S. pneumoniae antigens can also be used to  
 CC prevent or attenuate a Streptococcal infection in an animal. The  
 CC polynucleotides encoding the S. pneumoniae antigens can be used to  
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
 CC which are used in an example from the present invention.  
 XX  
 SQ Sequence 117 AA;  
 Query Match 77.0%; Score 615; DB 23; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-54;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 38 YKGELEKGYQFGWETSGFEGKDGAVINLSKDTFIKVPFKIEEKEENKPTFDVSK 97  
 DB 1 YKGELEKGYQFGWELSGFEGKDGAVINLSKDTFIKVPFKIEEKEENKPTFDVSK 60  
 QY 98 KKDNPQVNSQLNESHKEDLQREHSQKSDTKVATVLDKNNISSKSTTNNPNK 154  
 DB 61 KKDNPQVNSQLNESHKEDLQREHSQKSDTKVATVLDKNNISSKSTTNNPNK 117  
 RESULT 6  
 AAG81779  
 ID AAG81779 standard; Protein; 746 AA.  
 XX  
 AC AAG81779;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.  
 XX  
 DE Staphylococcus epidermidis SR1 strain; infection; diagnosis;  
 KW vaccination; endocarditis.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN WO200134809-A2.  
 XX  
 PD 17-MAY-2001.  
 XX  
 XX 09-NOV-2000; 2000WO-US30782.  
 PF  
 XX 09-NOV-1999; 99US-0164258.  
 PR  
 XX (GLAX ) GLAXO GROUP LTD.  
 PA  
 XX Kimmerly WJ;  
 PI  
 XX WPI; 2001-316495/33.  
 DR N-PSDB; AAH52629.  
 DR  
 XX

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 PS Claim 18; Page 208; 2188pp; English.  
 XX  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX  
 SQ Sequence 746 AA;  
 Query Match 14.8%; Score 118; DB 22; Length 746;  
 Best Local Similarity 27.2%; Pred. No. 0.0063;  
 Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
 QY 1 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYK-----GELEKGYQFGW--- 51  
 DB 586 DSVNAQSLKP-----ITIGNKQIKQOSVKSGTKVLPHSKVLMITDGLTMP-DMTGTWKE 640  
 QY 52 EISGFE-----GKDGAVIN--LSKDTFIKVPFKIEEKEENKPTFDVS-----KK 98  
 DB 641 DVLADEFDTKLKVSTKNGFVNTQSIKQGIK-----NKKDKIEVLSAEDT 687  
 QY 99 KDNPOVNSQLNESHKEDLQREHSQKSDTKVATVLDKNNISSKSTTNN 151  
 DB 688 DDDQKTDSDSNKSKKDADEHSNTSSSTKN-----DKSNADSKNDSDD 734  
 RESULT 7  
 ABP39023  
 ID ABP39023 standard; Protein; 778 AA.  
 XX  
 AC ABP39023;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.  
 XX  
 DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 PF 13-AUG-1998; 98US-0134001.  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX



```

DR WFI; 2002-391255/41.
DR N-PSDB; ABN91568.
XX
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX
PS Disclosure; SEQ ID 3868; 267pp; English.
XX
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
XX
SQ Sequence 778 AA;

      Query Match          14.8%; Score 118; DB 23; Length 778;
Best Local Similarity    27.2%; Pred.No. 0.0067;
Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 3

QY      1 DTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYK-----GELEKGQYFDGW--- 51
Db      618 DSVNAQLAP-----ITIGNGKIQKQS VKSGTKVLPHSKVLMLTDELTMP-DMTGWTKE 67
QY      52 EISGFE-----GKKDAGYVIN--LSKDTFTIKVPFKKIEKKEENKPTFDVS----KK 98
Db      673 DVLAPELDLKIKVSTKGNCFVTNQSIKSGLIHK-----NKDKIEVSLSAEDT 71
QY      99 KNPQVNHSQLNESHREKDLQREEHSQKSDSTKVDTATVLDKNWISSKSTNN 151
Db      720 DDQOEKTDEDSDNKSKOKKADEHSHNTSSSTKN-----DKSNADSKNDSD 766

RESULT 8
AAG47777
XX AAG47777 standard; Protein; 484 AA.
XX
XX AAG47777;
XX
XX 18-OCT-2000 (first entry)
DE
XX Arabidopsis thaliana protein fragment SEQ ID NO: 60255.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.

```



PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
 XX diagnosis of P.falciparum infection -  
 PS Disclosure; Page 321-322; 577pp; English.  
 XX  
 CC The present invention describes proteins and their fragments (I) encoded  
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
 CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)  
 CC vaccines against P. falciparum infection comprising (i) or (ii).  
 CC (i) and (ii) are useful for the development of vaccines against  
 CC P. falciparum infection. (i) and polyclonal antisera or a monoclonal  
 CC antibody raised to immunogens comprising the sequences of (i), are  
 CC useful in the detection of infection with P. falciparum. Furthermore,  
 CC (i) (especially when they are rifins or secreted or membrane proteins)  
 CC can aid the identification of drugs to treat or prevent P. falciparum  
 CC infection, or they can be used to identify drug resistance in  
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
 CC subsequent identification of proteins encoded by it will help to expand  
 CC our understanding of parasite biology, a process hampered by the  
 CC complexity of the parasitic lifecycle, and provide new targets for  
 CC vaccine and drug development. Parasite resistance to drugs and mosquito  
 CC resistance to insecticides have led to a resurgence of malaria in many  
 CC parts of the world, and there is a pressing need for vaccines and new  
 CC drugs. AAV0078 to AAA0287 and AAB18144 to AAB18352 represent nucleotide  
 CC and protein sequences given in the present invention, but which are not  
 CC specifically mentioned within the specification.  
 XX  
 SQ Sequence 665 AA;

Query Match 13.1%; Score 105; DB 21; Length 665;  
 Best Local Similarity 24.4%; Pred. No. 0.11;  
 Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;  
 Qy 1 DTGEVSELPKPRVT-VTTONKEMSSSTIVSEEDFILPVYKGELEKGYQFGWEI--SGFE 57  
 Db 128 EKNKINKSDLRQNELNLSQSGK-----NEQDI-----NKNEKGKQ-----DISNSNAE 170  
 Qy 58 GKQDAGYVINLSKDTFKIPVFKKIEEKE-----EENKPTFD-----VSKKKNQNP 103  
 Db 171 NKKD-----VKEGVKELEKKEKKEKISDDHKVEENKSDDDHKVEENKSDDHK 218  
 Qy 104 VNHSQLNESHKREDLQR-EHHSQKSDSKDTVTATVLDKRNISSKSTTNPNK 154  
 Db 219 VEENKSDDHKITEEVKVEEHEDEEE-----DKKEKSENKKNKDNENK 261

RESULT 10  
 ID ABB61977  
 AC ABB61977 standard; Protein; 564 AA.  
 AC ABB61977;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 12723.

KW Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL06080.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 FT genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX interactions -  
 PS Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 564 AA;

Query Match 12.7%; Score 101.5; DB 22; Length 564;  
 Best Local Similarity 24.5%; Pred. No. 0.19;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 Qy 31 EDFILPVYKGELEKGYQFGW-----EISGFEGKKDAGYVI-----NLSKDTFIK 75  
 Db 78 EDLPTLSERFSK--VPDGVVDEHRDHDGHVQEPSGEALDDHDDHDEHDEDEE 135  
 Qy 76 PVFKKIEEKEENKPT-----PDVSKKKNPQVNHSQLNESHKREDLQREHHSQKSDS 129  
 Db 136 PLTSEEEEEEPEETEDPEADEEVEDEENNA--GENITAEDEEEEDND 193  
 Qy 130 TKQVTATVLDKRNISSKST 148  
 Db 194 EGTVEATVEATTEATTEAT 212

RESULT 11  
 ID AAE20967  
 ID AAE20967 standard; Protein; 2060 AA.

AC AAE20967;

XX 01-JUL-2002 (first entry)

DT Staphylococcus lugdunensis von Willebrand factor binding protein.

DE Von Willebrand factor binding protein; vWb; immunogen; antibacterial;  
 KW vaccine; infection.

XX Staphylococcus lugdunensis.

XX Key Location/Qualifiers

FT Peptide 1..47

FT Protein /label= Signal\_peptide  
 48..2060

FT /note= "Mature von Willebrand factor binding protein"

XX WO200228892-A1.

XX 11-APR-2002.

XX 06-APR-2001; 2001WO-SE00766.

XX 04-OCT-2000; 2000SE-0003573.

XX (BIOS-) BIOTAPRO AB.

XX Guss B, Frykberg L, Jacobsson K, Ahlen J, Nilsson M;

```
XX WPI; 2002-304928/34.
DR N-PSDB; AAD33371.
XX
XX New von Willebrand factor binding protein from Staphylococci, useful
PT for determining and treating staphylococcal infection
XX
XX Claim 3; Page 35-41; 53pp; English.
XX
XX The present invention relates to von Willebrand factor binding protein or
CC polypeptide (vwb) from Staphylococci. The vwb and immunogens of vwb are
CC useful in vaccines to combat infections caused by Staphylococci. The
CC invention is also useful for detection of staphylococcal infection and
CC purifying von Willebrand factor from a complex solution. The present
CC sequence is Staphylococcus lugdunensis vwb1 protein.
XX
XX Sequence 2060 AA;
SQ
Query Match 12.1%; Score 96.5; DB 23; Length 2060;
Best Local Similarity 25.7%; Pred. No. 3.8; Mismatches 25; Gaps 9;
Matches 39; Conservative 25; Indels 49;
QY 4 EVSELKPHRVTVTIQ-NGKEMSSTIVSEEDFILPVYKGELEKGYFDG--WBISGFEGKK 60
DB 1914 DIELK---ITIVDTNGRE-----IVPSRKGLPP-EQFIGODWQYTGHK--- 1955
QY 61 DAGYVINLSKDTFIKPVFKKIEE-----KKEENKPTFDVSKKKNPQVNHSQLNESH 113
DB 1956 -----IEKDGITTVYKKVENAVPAKQLKTKHN--TQSESQFKHTPQVKQLVKVHN 2006
QY 114 RKE--DLQREHSQKSDSTKDTATVLDKNNI 143
DB 2007 VKEQRSIEKSEHTMHVSELPTGETANKGL 2038
XX
RESULT 12
AAB18272
ID AAB18272 standard; Protein; 2500 AA.
XX
XX AAB18272;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:129.
DE
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoicide; infection; insecticide.
XX
XX Plasmodium falciparum.
OS
XX WO200025728-A2.
FN
XX 11-MAY-2000.
PD
XX 05-NOV-1999; 99WO-US26796.
PF
XX 05-NOV-1998; 98US-0107131.
PR
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malaria parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P. falciparum infection
XX
XX Disclosure; Page 302-309; 577pp; English.
XX
```

```
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malaria parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAT70078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 2500 AA;
SQ
Query Match 12.0%; Score 96; DB 21; Length 2500;
Best Local Similarity 26.1%; Pred. No. 5.5;
Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;
QY 9 KPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYFDGWEISGFEGKDGAVINL 68
DB 2173 KPYKIT---ENKK-----NEGNEILKKYSTENEKKNYDKEQENCILDKDTQCNVNT 2223
QY 69 -----SKDTFIKPVFKKIEEENKPTFDVSKKKNPQVNHSQLNE----SHRKED 117
DB 2224 KEKNLNDNKKSPSNIKVKLEEEKSDDKRD---DKKNDNTRKKNLNDNKKSPSNIKVK 2280
QY 118 LQREHSQKSDSTKDTATVLDKNNISSKSTTNNPNK 154
DB 2281 LEESE---KSDDKRD-----DKKNDNTRKKNLNDNKKSPSNIKVK 2308
XX
RESULT 13
AAG37134
ID AAG37134 standard; Protein; 408 AA.
XX
XX AAG37134;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 45610.
DE
XX Arabidopsis thaliana.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
FN
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
```

PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.

```
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.7%; Score 93.5; DB 21; Length 408;
Best Local Similarity 25.0%; Pred. No. 0.79;
Matches 39; Conservative 28; Mismatches 58; Indels 31; Gaps 8;

QY 4 EYSEKLP-HRVVTVTQNGKEMSTIVSEEDFILPVYKGEKGYQFDGWEISGFEKKDA 62
DB 201 EKDPLKPHPSAFILVYANERRAALREENKSVVEVAK-----ITGEWNKNSDKKA 252

QY 63 GY--VINLSKDTFIKPVFKKIBEKKEENKPTFDVSKKNDPNQVNSQLNESHKEDLQR 120
DB 253 PYEKVAKKNKEIYQ-AMEYKTRKEE-----ALSOKE-----EEELLKLHKQEAQOM 301

QY 121 EHSQKSDSTKQVATVLVDKNNISSKSTNN--PNK 154
DB 302 LKKKERTDN-----LIKKERATKKKNENVDPNK 330

RESULT 14
ID AAG37133 standard; Protein; 456 AA.
XX
AC AAG37133;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45609.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
```

```
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155133.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.7%; Score 93.5; DB 21; Length 456;
Best Local Similarity 25.0%; Pred. No. 0.92;
Matches 39; Conservative 28; Mismatches 58; Indels 31; Gaps 8;

Qy 4 EVSELKP-HRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDA 62
Db 249 EKDPLKPKHPVSAFLVTVANERRAALREENKSVVEVAK-----ITGEWKNLSDKKKA 300

Qy 63 GY--VINLSKDTFKIPVKKIBKKKEENKPTFDVSKKKNQPNVHSQLNESHKEDLQR 120
Db 301 PYEKVAKKNKETYLIQ-AMEEYKRTKEE-----ALSQKE-----BEELKLHKQEAQLQM 349

Qy 121 EHSOKSDSKDVTATVLDKNNISSKSTTN--PNK 154
Db 350 LKKKERTDN-----LIKKEKATYKKKNENVDNPK 378

RESULT 15
AAG37132
ID AAG37132 standard; Protein; 476 AA.
XX
AC AAG37132;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45608.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
FN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
```

PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.



PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 11.7%; Score 93.5; DB 21; Length 476;  
Best Local Similarity 25.0%; Pred. No. 0.98; Mismatches 28; Conservative 39; Gaps 8;  
Matches 39; Indels 31; Gaps 8;

QY 4 EVSELPK-HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFRGKKA 62  
DB 269 EKDPKPKHPVSAFLVYANERAAALREENKSVVEAK-----ITGEWKNLSDKKA 320  
QY 63 GY--VINLSKDTFIKPVFKIEEKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120  
DB 321 PYEKVAKKNKETYLO-AMBEYKRTKEE-----ALSQKE-----EEELLKLHKQEAQM 369  
QY 121 EHSQKSDSTKDVATVLDKKNISKSSTNN--PNK 154  
DB 370 LKKEKTDN-----LIKKEKATKKKNENVDPNK 398

RESULT 16  
ABG16636  
ID ABG16636 standard; Protein; 2519 AA.  
XX  
AC ABG16636;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #16627.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
FN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-0508631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DB WPI; 2001-639362/73.

DR N-PSDB; AAS80823.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 46995; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2519 AA;  
Query Match 11.6%; Score 92.5; DB 22; Length 2519;  
Best Local Similarity 31.0%; Pred. No. 13;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
QY 68 LSKDTFIKPVFKIEEKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR-----E 122  
DB 689 VKKETKVP-----EDKKEEKEKPKKEVAKEDKTPI---KKEEKPKEEKKVKEIK 740  
QY 123 HSQKSDSTKDV 133  
DB 741 KEEKKEPKKEV 751  
RESULT 17  
AAB42897  
ID AAB42897 standard; Protein; 209 AA.  
XX  
AC AAB42897;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF2661 polypeptide sequence SEQ ID NO:5322.  
XX  
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antitumour; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; erythematoidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
OS Homo sapiens.  
XX

PN WO200058473-A2.  
XX 05-OCT-2000.  
XX 31-MAR-2000; 2000WO-US08621.  
XX 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX (CURA-) CURAGEN CORP.  
XX Shimkets RA, Leach M;  
XX WPI; 2000-602362/57.  
XX N-PSDB; AAC77106.  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX Claim 11; Page 4497-4498; 5507pp; English.  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;  
CC osteoparic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX Sequence 209 AA;  
SQ  
Query Match 11.3%; Score 90.5; DB 21; Length 209;  
Best Local Similarity 28.1%; Pred. No. 0.62;  
Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;  
QY 3 GVSSELPKPHRVTVTQNGKEMSTIVSE--EDFILPVYKGELEKGYQF-----DGWE--- 52  
Db 18 GELQLUGSQL-----HGKSDSPNVTEKKEITALLRERUTELERKUTFTQQRSDDLWERLY 72  
QY 53 -----ISGPEGKKDAG-----YVINLSKDTFKPKFKKIBKKEENKPTFDVSKKID 100  
Db 73 VEAKDQNGKQGTGDKKKGGSHRAKNKSKFVLSGV-----KEITFAMKNST 120  
QY 101 NPOVNH-----SQLNESHKEDLQREHSQKSD--STKDVATVLDK 140  
Db 121 KEFVRHHEKIKQAKA--VKNLKKFSDSVKSTFRHFKDTTKNIFDE 166  
RESULT 18  
AA44364  
ID AAY44364 standard; protein; 757 AA.  
XX AAY44364;  
AC AAY44364;  
XX 14-MAR-2000 (first entry)  
DT  
XX

DE Human cell cycle regulation protein-5.  
XX  
KW CCRP-5; cell cycle regulation protein-5; cell proliferation;  
KW cell proliferative disease; cancer; atherosclerosis; cirrhosis;  
KW hepatitis; psoriasis; immune system disorder; allergy; asthma;  
KW acquired immune deficiency syndrome; Crohn's disease; Blast method;  
KW rheumatoid arthritis; gene therapy; chromosomal mapping.  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 4  
FT Modified-site /note= "Potential glycosylation site"  
FT Modified-site 63  
FT Modified-site /note= "Potential glycosylation site"  
FT Modified-site 479  
FT Modified-site /note= "Potential glycosylation site"  
FT Modified-site 498  
FT Modified-site /note= "Potential glycosylation site"  
FT Modified-site 31  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 43  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 81  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 96  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 99  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 130  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 158  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 162  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 163  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 184  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 186  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 188  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 271  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 277  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 287  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 289  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 323  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 343  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 367  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 370  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 432  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 488  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 499  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 207  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 277  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 298  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 370  
FT Modified-site /note= "Potential phosphorylation site"

FT	Modified-site	419	/note=	"Potential phosphorylation site"
FT	Modified-site	474	/note=	"Potential phosphorylation site"
FT	Modified-site	488	/note=	"Potential phosphorylation site"
FT	Modified-site	499	/note=	"Potential phosphorylation site"
FT	Modified-site	528	/note=	"Potential phosphorylation site"
FT	Modified-site	532	/note=	"Potential phosphorylation site"
FT	Modified-site	539	/note=	"Potential phosphorylation site"
FT	Modified-site	609	/note=	"Potential phosphorylation site"
FT	Modified-site	624	/note=	"Potential phosphorylation site"
FT	Modified-site	730	/note=	"Potential phosphorylation site"
FT	Modified-site		/note=	"Potential phosphorylation site"

XX PN WO9964593-A2

XX  
PD  
16-DEC-1999.

AA 08-JUN-1999: 99WO-US12906.

XX 08-JUN-1998: 98US-0088695.

PA (INCY-) INCYTE PHARM INC.

XX Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR;  
PI Patterson C;

WPI: 2000-105887/09.

DR WFI, 2000-103087/  
DR N-PSDB: AAZ29484.

Novel regulatory proteins, for diagnosis, treatment and prevention of cell proliferative and immune system diseases -

PS Claim 1; Page 70-71; 87pp; English.

The present sequence is cell cycle regulation protein-5 (CECRP-5). Blast analytical method was used to identify this protein. CECRPs are activators of cell proliferation or inhibitors of cellular processes that modulate proliferation. They are used to treat or prevent cell proliferative diseases like cancers, atherosclerosis, cirrhosis, hepatitis, psoriasis, immune system disorders (e.g. acquired immune deficiency syndrome, allergy, asthma, Crohn's disease, rheumatoid arthritis). Antibodies are raised to screen for specific binding agents. The corresponding nucleic acid is used in gene therapy, chromosomal mapping and isolation of related sequences.

XX	Sequence	757 AA:
S0		

Query Match	11.3%	Score 90.5;	DB 21;	Length 757;
Best Local Similarity	28.1%	Pred. No. 3.7;		
Matches	47: Conservative	17: Mismatches	56: Indels	47: Deletions

[illegible]

QY	53	-----ISGFEGKQDAG-----YVINLSKDTFIKVPFKKIEKKKEENKPTTFDVS	KKD	100
DP	453	VEAKDONGKQGTGDKKKGGRGSHRAKNKSKETFLGSV-----KETFDA	MKNST	500

101 NPOVNH-----SOLNESHKEDLOREEHSOKSD--STKDVTAATVLPK 140

501 KEFVRHHKEKI KOAKEA - VKENLKFSDSVKSTFRHFKOTTTQNI FEDE 546

## RESULT 19

AAB53319

ID AAB53319 standard; Protein; 758 AA.

AC AAB53319;

09-MAR-2001 (first entry)

xx DE Human colon cancer antigen protein sequence SEQ ID NO:859.

Human; colon cancer; colon cancer antigen; diagnosis; detection; identification; cytostatic; cardioactive; neuroprotective; vulnary; immunomodulatory; muscular; gynaecological; gastrointestinal; nephrotropic; antibacterial; gene therapy; wound; neural disorder; immune system disorder; muscular disorder; reproductive disorder; gastrointestinal disorder; renal disorder; infectious disease; cardiovascular disorder.

XX Homo sapiens.

XX PN WO200055351-A1.

XX  
PD 21-SEP-2000.

08-MAR-2000: 2000WO-US05883.

XX  
PR 12-MAR-1999: 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX  
PI  
Rosen CA. Ruben SM:

XX  
DB  
WPT: 2000-587534/55.

DR N-PSDB; AAC98076.  
XX

PT antigens, useful for the treatment, prevention, and diagnosis of colon  
PT disorders such as colon cancer -  
PT Colonic cancer associated gene sequences, referred to as colon cancer  
PT Colonic cancer associated gene sequences, referred to as colon cancer

PS Claim 11: Page 1408-1411; 2104pp; English.

AAC97991 to AAC98763 encode the human colon cancer associated proteins, called human colon cancer antigens, given in AAB53234 to AAB54006. The human colon cancer antigens can have cytostatic, cardioactive, muscular; neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and vulvular, nephrotropic, antineoplastic and antibacterial activities, and can be used in gene therapy. The colon cancer antigen polynucleotides, proteins and antibodies to the proteins are useful for the prevention, treatment and diagnosis of colon disorders, such as colon cancer. The polynucleotides may be used in diagnostics and research, such as for chromosome identification, and as hybridisation probes. The proteins may also be used to prevent diseases such as neural disorders, immune system disorders, muscular disorders, reproductive disorders, gastrointestinal disorders, wounds, renal disorders, infectious diseases, and cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent sequences used in the exemplification of the present invention.

XX  
50 Sequence 758 AA:

Query Match	11.3%	Score 90.5;	DB 21;	Length 758;
Best Local Similarity	28.1%;	Pred. No. 3.8;		
Matches	47: Conservative	17: Mismatch	56: Indels	47: Gaps

Qy 3 GEVSELPKPRVTVTIQNGKMSSTIVSE--EDFILPVYKGELEKGYQF-----DGWE--- 52

Dh 399 GETLOOLSGSOL-----HGKSDSPNVYTEKKKIAILRELTTEPKLTPCOORSPLMERLY 45

53 -----TSGFEGKKDAG-----YVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKD 100

db 454 VEAKDONGKOGTDGKKKGGRGSHRAKNKSKETFLGSV-----KETFDAKNST 501

QY 101 NPQVNH-----SQLNESHKEDLQREHSQKSD--STKDVTTATVLDK 140

Db 502 KSFVRHKEKIKQAKEA-VKENLKPFSDSVKSTFRHFKDTTKNIFDE 547

```
RESULT 20
ABP38188
ID ABP38188 standard; Protein; 442 AA.
XX AC ABP38188;
XX DT 24-JUL-2002 (first entry)
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3033.
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX KW antibacterial; gene therapy.
XX OS Staphylococcus epidermidis.
XX PN US6380370-B1.
XX PD 30-APR-2002.
XX PF 13-AUG-1998; 98US-0134001.
XX PR 14-AUG-1997; 97US-055779P.
XX PR 08-NOV-1997; 97US-064964P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm LA, Bush D;
XX PI WPI; 2002-381255/41.
XX DR N-PSDB; ABN90733.
XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX PT polypeptide, useful for diagnosing and treating bacterial infections -
XX PS Disclosure; SEQ ID 3033; 267pp; English.
XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX CC antibacterial activity and can be used in gene therapy. The sequences
XX CC can also be used in the diagnosis and treatment of bacterial infections,
XX CC particularly S. epidermidis infections. The sequences can be used to
XX CC screen for compounds able to interfere with the S. epidermidis life
XX CC cycle or inhibit S. epidermidis infection.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC USPTO web site.
XX SQ Sequence 442 AA;
Query Match 11.3%; Score 90; DB 23; Length 442;
Best Local Similarity 21.8%; Pred. No. 2;
Matches 49; Conservative 29; Mismatches 65; Indels 82; Gaps 11;
QY 1 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPYKG-ELEK----- 44
Db 91 DT-EINPVAQPEVNVTVQIEKGRDFEATVVEPEVKLGDKYKLEIEKETDLSDEELQE 149
QY 45 -----GYQFDGWEISG--FEGKKDAGYVINLSKDTFIKP 76
Db 150 SIDHLSHLAEMVMVKEDGAVENGDTVNIDFSG-SVDGEFDDGQAGYDLIGSGSFIPG 208
QY 77 VFKKIEEKEENK-----PTFDVS-----KKDNQVNHSQLNE-- 111
Db 209 FEEQIEGMKTGDKDQVVVTFPEYHAEELAGKATFKTKNEIKFKDVPDELNDIANELD 268
QY 112 -----SHRKEDQREEHSSQKSDSTKDVATATVLDKNNISSKSTTN 150
Db 269 SDAENVDEYKENLRKLSQKATEAENT-----EKEEAINKATEN 308
```

RESULT 21

```
ABU19106
ID ABU19106 standard; Protein; 645 AA.
XX AC ABU19106;
XX DT 06-MAR-2003 (first entry)
XX DE Pathogen specific antigen related staphylococcal protein SEQ ID No 414.
XX KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
XX KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
XX KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
XX KW autoimmune disease; HIV; hepatitis.
XX OS Staphylococcus sp.
XX PN WO200259148-A2.
XX PD 01-AUG-2002.
XX PF 21-JAN-2002; 2002WO-EP00546.
XX PR 26-JAN-2001; 2001AT-0000130.
XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX PI Meinke A, Nagy E, Von Ahseu U, Klade C, Henics T, Zauner W;
XX PI Minke DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;
XX PI Tempelmaier B;
XX DR WPI; 2003-075410/07.
XX PT Identifying, isolating and producing hyperimmune serum-reactive
XX PT antigens from a pathogen, for preparing vaccine or medicament for
XX PT treating or preventing e.g. staphylococcal infections, comprises
XX PT providing antibody preparation -
XX PS Example 7; Page 220; 252pp; English.
XX CC The invention relates to a novel method for identifying, isolating and
XX CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
XX CC allergen, a tissue or host prone to auto-immunity, where the antigens
XX CC are used in a vaccine, comprises providing antibody preparation from a
XX CC plasma pool of a type of animal, or individual sera with antibodies
XX CC against the specific pathogen, tumour, allergen, tissue or host prone to
XX CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
XX CC the 62 sequences of 53-2261 amino acids fully defined in the
XX CC specification, or their hyperimmune fragments are useful for the
XX CC manufacture of a pharmaceutical preparation, particularly a vaccine
XX CC against staphylococcal infections or colonisation against S. aureus or S.
XX CC epidermidis. The preparation of antibodies is useful for the manufacture
XX CC of a medicament for treating or preventing staphylococcal infections or
XX CC colonisation against S. aureus or S. epidermidis. The antibody
XX CC preparations may also be used for diagnostic and imaging purposes. Other
XX CC conditions that can be treated include cancer, autoimmune diseases or
XX CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
XX CC protozoan pathogens. This sequence represents a staphylococcal protein
XX CC relating to the method for identifying and producing pathogen specific
XX CC antigens of the invention.
XX SQ Sequence 645 AA;
Query Match 11.3%; Score 90; DB 24; Length 645;
Best Local Similarity 21.4%; Pred. No. 3.4;
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;
QY 4 EYSELKPHRVTV--TIQNGKMSSTIVSEEDFILPYKGELB-KGYQF-----DGWELS 54
Db 341 KMTDLQDTKYVYVESVENNESMMDTFVKH-----PIKTGMLNGKKYVMVMTTNDYKDF 395
QY 55 GFEGKK-----DAGYVINL-SKDTFI 74
```

Db 396 MVEGQVVRTISKDAKNNTRTIIPYVEGKTLDAIVKVHVKTIDYDGYHVRIVDKAFT 455  
Qy 75 KPVFKKIEEKEENKPTFDV-----SKKDNQPNVHSQLNESHKEDLQ-----REHSQ 125  
Db 456 KANTDKSNKQKQDQNSAKKEATPATPSKPTSPVKEKSKQKQDQDNKQLPSVEKENDA 515  
Qy 126 KSDSTKQVT-ATVLDKNNISSKSTTNNPNK 154  
Db 516 SSESCKDTPATKPTKGEVSSSTT--PTK 543

## RESULT 22

ABP56879  
ID ABP56879 standard; Protein; 654 AA.

XX AC ABP56879;

XX DT 07-APR-2003 (first entry)

XX DE Staphylococcus epidermidis KrKN protein SEQ ID NO:10.

XX KW Staphylococcus aureus; Staphylococcus epidermidis; MSCRAMM; antibody;  
KW microbial surface component recognising adhesive matrix molecule;  
KW surface protein; infection; antibacterial; antiinflammatory; vaccine;  
KW immunosuppressive; antiarthritic; gene therapy; pneumonia; endocarditis;  
KW septic arthritis; biomaterial related infection.

XX OS Staphylococcus epidermidis.

XX PN WO2002102829-A2.

XX PD 27-DEC-2002.

XX PF 17-JUN-2002; 2002WO-US19220.

XX PR 15-JUN-2001; 2001US-298098P.

XX PA (INHI-) INHIBITEX INC.

XX PA (QUEE-) QUEEN ELIZABETH COLLEGE DUBLIN.

XX PA (UTPA-) UNIV PAVIA.

XX PI Foster TJ, Roche F, Patti JM, Hutchins JT, Hall A, Domanski P;  
XX Patel P, Syribeys P, Speziale P;

XX WPI; 2003-167481/16.

XX DR N-PSDB; ABZ22903.

XX PT New isolated antibody that binds to a staphylococcal surface protein,  
XX useful for treating or preventing Staphylococcus aureus infections,  
XX such as pneumonia, septic arthritis, endocarditis or biomaterial  
XX related infections -

XX PS Claim 1; Page 29-30; 122pp; English.

XX CC The present invention describes an isolated antibody (I) that binds to a  
XX staphylococcal surface protein selected from any of the 12 sequences of  
XX 354-2283 amino acids given in ABP56875 to ABP56886. Also described:  
XX (1) an isolated antisera (II) comprising (I); (2) a diagnostic kit (III)  
XX comprising (I) and means for detecting binding by (I); (3) diagnosing  
XX (MI) an infection of Staphylococcus aureus by adding (I) to a sample  
XX suspected of being infected with the infection, and determining if  
XX antibodies have bound to the sample; (4) a pharmaceutical composition  
XX (IV) for treating or preventing an infection of S. aureus comprising (I),  
XX and a vehicle, carrier or excipient; (5) treating (M2) or preventing an  
XX infection of S. aureus by administering (I) to a human or animal patient;  
XX (6) producing (M3) an immunological response by administering to a human  
XX or animal an immunogenic amount of the isolated surface protein; (7) an  
XX isolated active fragment (V) from the A domain of the DsA protein; and  
XX (8) a vaccine (VI) for treating or preventing an infection of S. aureus  
XX comprising the surface protein in an amount effective to elicit an immune  
XX response, and a vehicle, carrier or excipient. (I) has antiinflammatory,  
XX antibacterial, immunosuppressive and antiarthritic activities, and can be

CC used in gene therapy. The antibody, composition and vaccine are useful  
CC for treating or preventing Staphylococcus aureus infection in a human or  
CC animal, such as pneumonia, septic arthritis, endocarditis or biomaterial  
CC related infections. The present sequence represents Staphylococcus  
CC epidermidis KrKN protein, which is used in the exemplification of  
XX the present invention.

XX SQ Sequence 654 AA;

Query Match 11.3%; Score 90; DB 24; Length 654;  
Best Local Similarity 21.4%; Pred. No. 3.4;  
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;

Qy 4 EVSELKPHRTV--TIQNKEMSTIVSEEDFILPVYKGELE-KGYQF-----DGWEIS 54

Db 350 KMTDLQDTKYVYVESVENNESMDTFFVKH-----PIKTGMLNGKKYVMYMETTNDYWKDF 404

Qy 55 GFEGKK-----DAGYVNL-SKOTFI 74

Db 405 MVEGQVVRTISKDAKNNTRTIIPYVEGKTLDAIVKVHVKTIDYDGYHVRIVDKAFT 464

Qy 75 KPVFKKIEEKEENKPTFDV-----SKKDNQPNVHSQLNESHKEDLQ-----REHSQ 125

Db 465 KANTDKSNKQKQDQNSAKKEATPATPSKPTSPVKEKSKQKQDQDNKQLPSVEKENDA 524

Qy 126 KSDSTKQVT-ATVLDKNNISSKSTTNNPNK 154

Db 525 SSESCKDTPATKPTKGEVSSSTT--PTK 552

## RESULT 23

ABBS8769

ID ABBS8769 standard; Protein; 1183 AA.

XX AC ABBS8769;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 3099.

XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL02872.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -

XX PS Disclosure; SEQ ID NO 3099; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB57737-AB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1183 AA;

Query Match 11.3%; Score 90; DB 22; Length 1183;  
 Best Local Similarity 24.7%; Pred. No. 7.8;  
 Matches 36; Conservative 24; Mismatches 54; Indels 32; Gaps 5;

QY 9 KHRVTVTQNGKMSSTIVSEDFILPVYKGELEKGYQDFGWEISGFGKDGAVINL 68  
 DB 310 KPEVKTAKDNKDD---PLIADSKPKVPKPKTSK-----TGGKKKEGA----- 352

QY 69 SKDTFTKPVFKTEEEKKEENKPTFDVSKKDNQPNVHSQLNESHKEDLOREHSQ--- 125  
 DB 353 -----KPAEK--SKEEKEDSTTKSKKKAUSPANNOKEQIQGVKKTSSEPDASNSHK 404

QY 126 ----KSDSTKDVATVLDKNNISSKS 147  
 DB 405 SVTAKKSKKDEASTQVKSDNESPEA 430

RESULT 24  
 ID AAY03190 standard; Protein; 402 AA.  
 AC AAY03190;  
 XX  
 DT 16-JUN-1999 (first entry)  
 DE  
 DE S. aureus trigger factor protein sequence.  
 KW Trigger factor; tig; diagnosis; immune response; bacterial infection;  
 KW S. aureus infection; otitis media; toxic shock syndrome; conjunctivitis;  
 KW wound infection; impetigo; septic arthritis; therapy.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PN EP905233-A2.  
 XX  
 PD 31-MAR-1999.  
 XX  
 PF 21-AUG-1998; 98BP-0306697.  
 XX  
 PR 29-DEC-1997; 97US-0999339.  
 PR 04-SEP-1997; 97US-0057511.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Burnham MKR, Fosberry A, Hodgson JE, Jaworski DD;  
 PI Lawlor EJ, Rosenberg M, Trainl CM, Wang M, Ward J;  
 PI Warren RL;  
 XX  
 DR WPI; 1999-192663/17.  
 DR N-PSDB; AAX28175.  
 XX  
 PT New Staphylococcus aureus trigger factor (tig) polypeptide and  
 PT polynucleotide, useful as diagnostic reagents and for prevention and  
 PT treatment of Staphylococcus aureus infections  
 XX  
 XX Claim 23; Page 8; 31pp; English.  
 XX  
 CC This sequence is the Staphylococcus aureus trigger factor (tig)  
 CC polypeptide of the invention. Tig polypeptides and polynucleotides are  
 CC useful for diagnosing diseases related to over or underexpression of tig  
 CC protein by identifying mutations in the tig gene, or determining tig  
 CC polypeptide or mRNA expression levels due to an infection of an organism  
 CC with the tig gene. They can diagnose the stage and type of infection. Tig  
 CC polypeptides are also useful for screening compounds which affect

CC activity of the protein by measuring the binding to tig and observing the  
 CC stimulation or inhibition of the polypeptide function. These can be used  
 CC in treatment to inhibit or enhance tig activity, in addition to direct  
 CC administration of tig polypeptides to treat conditions associated with a  
 CC lack of tig polypeptide, or direct administration of antisense sequences  
 CC to prevent expression. Tig polypeptides (administered directly, in a  
 CC vector and as a vaccine) and antibodies induce an immune response to  
 CC immunise and prevent disease. Diseases diagnosed, prevented or treated  
 CC include: bacterial infections, especially Staphylococcus aureus  
 CC infections which cause otitis media, toxic shock syndrome, Tig  
 CC conjunctivitis, wound infection, impetigo and septic arthritis etc. Tig  
 CC polypeptides, polynucleotides and their (ant)agonists can prevent  
 CC adhesion of bacteria to matrix proteins, and are useful for use on wounds  
 CC and body implants to prevent bacterial infection.  
 XX  
 SQ Sequence 402 AA;

Query Match 11.2%; Score 89.5; DB 20; Length 402;  
 Best Local Similarity 22.1%; Pred. No. 2;  
 Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 1 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 44  
 DB 53 DETDIKPVQAQPEVSVTQIEKGKDFEATVTVEPVKLGDKGLBIEKQETELSDDELQE 112

QY 45 -----GYQFDGWEISG--FEGKDGAGVIVNLSKDTFIKP 76  
 DB 113 AIDHSLGHLAEMVVKEDGVVNGDVTNIDFSG-SVDGSEFEGGQAEYDLEIGSGSFI-P 170

QY 77 VFK-----KTEEKKE-----EE--NKPTFDVS-----KKDNQPNVHSQLN- 111  
 DB 171 GFEEQLEGKMDDEKDVVTPPEYHABELAGKATFKTKVNEIKFKEVPELTDEIANEL 230

QY 112 -----SHRKEDLOREHSQKSDTKDVATVLDKNNISSKSTTN 150  
 DB 231 DAENTVDEYKENLKRKLAEGKATDAENV-----EKEBAITKATDN 271

RESULT 25  
 ID AAY35091 standard; Protein; 511 AA.  
 XX  
 AC AAY35091;  
 XX  
 DT 13-SEP-1999 (first entry)  
 DE  
 DE Chlamydia pneumoniae transmembrane protein sequence.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO9927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 XX (GEST ) GENSET.  
 XX  
 XX Griffais R;  
 XX WPI; 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 XX  
 PS Page 975-976; Disclosure; 1912pp; English.  
 XX

CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAX91990) of *Chlamydia pneumoniae*.  
 CC *C. pneumoniae* causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the *C. pneumoniae* genome (see AAY34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing *C. pneumoniae*  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of *C. pneumoniae*.  
 CC  
 XX Sequence 511 AA;  
 SQ  
 Query Match 11.2%; Score 89.5; DB 20; Length 511;  
 Best Local Similarity 24.5%; Pred. No. 2.7;  
 Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
 QY 74 IKPVFKIEKKEENKPTFD-----VSKKDNQVNHSQLNESHK 115  
 Db 95 VKGVFKTPQARPEVSSPLPSHVQHQRPLGLEGRDRIOKRSNPEDLGKMKRSYSD 154  
 QY 116 EDLQREHSQKSDTKDVTATVLDKNNISSKSTT 149  
 Db 155 GDLRVGHDSNEDSTEDSRS---EGGEPSSKSS 185  
 RESULT 26  
 ID AAY03189  
 AC AAY03189 standard; Protein; 525 AA.  
 XX  
 XX AAY03189;  
 DT 16-JUN-1999 (first entry)  
 DE  
 DE S. aureus trigger factor protein sequence.  
 KW Trigger factor; tig; diagnosis; immune response; bacterial infection;  
 KW S. aureus infection; otitis media; toxic shock syndrome; conjunctivitis;  
 KW wound infection; impetigo; septic arthritis; therapy.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 XX EP905233-A2.  
 XX  
 XX 31-MAR-1999.  
 XX  
 XX 21-AUG-1998; 98EP-0306697.  
 XX  
 XX 29-DEC-1997; 97US-0999339.  
 PR  
 PR 04-SEP-1997; 97US-0057511.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 XX Burnham MKB, Fosberry A, Hodgson JE, Jaworski DD;  
 PI Lawlor EJ, Rosenberg M, Traini CM, Wang M, Ward J;  
 PI Warren RL;  
 XX  
 XX WPI; 1999-192663/17.  
 DR  
 DR N-PSDB; AAX28174.  
 XX  
 XX New Staphylococcus aureus trigger factor (tig) polypeptide and  
 PT polynucleotide, useful as diagnostic reagents and for prevention and  
 PT treatment of Staphylococcus aureus infections  
 XX  
 XX Claim 14; Page 6; 31pp; English.  
 PS  
 PS This sequence is the Staphylococcus aureus trigger factor (tig)  
 CC polypeptide of the invention. Tig polypeptides and polynucleotides are  
 CC useful for diagnosing diseases related to over or underexpression of tig  
 CC protein by identifying mutations in the tig gene, or determining tig  
 CC polypeptide or mRNA expression levels due to an infection of an organism

CC with the tig gene. They can diagnose the stage and type of infection. Tig  
 CC polypeptides are also useful for screening compounds which affect  
 CC activity of the protein by measuring the binding to tig and observing the  
 CC stimulation or inhibition of the polypeptide function. These can be used  
 CC in treatment to inhibit or enhance tig activity, in addition to direct  
 CC administration of tig polypeptides to treat conditions associated with a  
 CC lack of tig polypeptide, or direct administration of antisense sequences  
 CC to prevent expression. Tig polypeptides (administered directly, in a  
 CC vector and as a vaccine) and antibodies induce an immune response to  
 CC immunise and prevent disease. Diseases diagnosed, prevented or treated  
 CC include: bacterial infections, especially Staphylococcus aureus  
 CC infections which cause otitis media, toxic shock syndrome,  
 CC conjunctivitis, wound infection, impetigo and septic arthritis etc. Tig  
 CC polypeptides, polynucleotides and their (ant)agonists can prevent  
 CC adhesion of bacteria to matrix proteins, and are useful for use on wounds  
 CC and body implants to prevent bacterial infection.  
 XX  
 XX Sequence 525 AA;  
 SQ  
 Query Match 11.2%; Score 89.5; DB 20; Length 525;  
 Best Local Similarity 22.1%; Pred. No. 2.8;  
 Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;  
 QY 1 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEDFILPVYKG-ELEK----- 44  
 Db 81 DETDIKPVAPQPEVSVTQIEKGDFIFEATVTVPEVKLGDKGLEIEKQETELSDDELQE 140  
 QY 45 -----GVQPDGWEISG--PEGKDGAGVIVNLSKDTFKP 76  
 Db 141 AIDHSLGHIAEMVVKEDGVENGDTVNIIDFSG-SVDGEFEGGQAEVDLISGSF1-P 198  
 QY 77 VFK-----KIEEKK-----EE--NKPTFDVS-----KKKDNQVNHSQLNE- 111  
 Db 199 GFEEQLGEMKVDDEKDVVVTPPEYHAEELAGKATFKTKVNEIKFKEVPDLTDEIANEL 258  
 QY 112 -----SHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTN 150  
 Db 259 DAEANTVDEYKENLRKRLAEQKATDAENV-----EKEEAITKATDN 299  
 RESULT 27  
 ABB58704  
 ID ABB58704 standard; Protein; 1408 AA.  
 XX  
 XX ABB58704;  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT  
 DE Drosophila melanogaster polypeptide SEQ ID NO 2904.  
 XX  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 XX Drosophila melanogaster.  
 OS  
 XX WO200171042-A2.  
 PN  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX 23-MAR-2000; 2000US-191637P.  
 PR  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 PI WPI; 2001-656860/75.  
 DR  
 DR N-PSDB; ABL02807.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell

```
PT interactions -
XX PT Disclosure; SEQ ID NO 2904; 2lpp + Sequence Listing; English.
XX PS
XX SS Claim 21; Page 168; 252pp; English.
XX CC The invention relates to a novel method for identifying, isolating and
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SS Sequence 1408 AA;
Query Match 11.2%; Score 89.5; DB 22; Length 1408;
Best Local Similarity 24.7%; Pred. No. 11;
Matches 48; Conservative 23; Mismatches 62; Indels 61; Gaps 9;
QY 19 NGKE-----MSSITVSEDF--ILPVYKGELEKGYQPDGWEISGFEGK--KDAGYV----- 65
Db 646 NGKQKLNASKEYTKSEKDFEIVESKKE-ERNDNQESKELENQDGKLPKQDYLQVKKP 704
QY 66 -----INLSKDTFKPVFK-----KIEKKEEENKPTFDVSKKKNQPNQVNHSQLN 110
Db 705 MESKEYSEPLAVKLGEDVSKSPNESSDTKISEVKGEE--PKMNGELPKTGQVKQPEKS 762
QY 111 ESHRK-EDIQREHSOKSPSTKQDVT-----ATVLQK 140
Db 763 KTOAKTEFLAKQSELSKSVKXDVTPEQTTKEETHKGFSPKREATPKSEBPKGTTEVDK 822
QY 141 NNISKSSTNNPNK 154
Db 823 SNDVSVIKSEPDE 836
RESULT 28
ID ABJ18979 standard; Protein; 645 AA.
XX AC ABJ18979;
XX DT 06-MAR-2003 (first entry)
XX DE Pathogen specific antigen related staphylococcal protein SEQ ID No 142.
XX KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
XX KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
XX KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
XX KW autoimmune disease; HIV; hepatitis.
XX OS Staphylococcus sp.
XX PN WO200259148-A2.
XX PD 01-AUG-2002.
XX PF 21-JAN-2002; 2002WO-EP00546.
XX PR 26-JAN-2001; 2001AT-0000130.
XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX PI Meinke A, Nagy E, Von Ahsen U, Klade C, Henics T, Zauner W;
XX PI Minn DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;
XX PI Tempelmaier B;
XX DR WPI; 2003-075410/07.
XX PT Identifying, isolating and producing hyperimmune serum-reactive
PT antigens from a pathogen, for preparing vaccine or medicament for
PT treating or preventing e.g. staphylococcal infections, comprises
PT providing antibody preparation -
XX SS Claim 21; Page 168; 252pp; English.
XX CC The invention relates to a novel method for identifying, isolating and
XX producing hyperimmune serum-reactive antigens from a pathogen, tumour,
XX allergen, a tissue or host prone to auto-immunity, where the antigens
XX are used in a vaccine, comprises providing antibody preparation from a
XX plasma pool of a type of animal, or individual sera with antibodies
XX against the specific pathogen, tumour, allergen, tissue or host prone to
XX auto-immunity. The hyperimmune serum-reactive antigens comprising any of
XX the 62 sequences of 53-2261 amino acids fully defined in the
XX specification, or their hyperimmune fragments are useful for the
XX manufacture of a pharmaceutical preparation, particularly a vaccine
XX against staphylococcal infections or colonisation against S. aureus or S.
XX epidermidis. The preparation of antibodies is useful for the manufacture
XX of a medicament for treating or preventing staphylococcal infections or
XX colonisation against S. aureus or S. epidermidis. The antibody
XX preparations may also be used for diagnostic and imaging purposes. Other
XX conditions that can be treated include cancer, autoimmune diseases or
XX infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
XX protozoan pathogens. This sequence represents a staphylococcal protein
XX relating to the method for identifying and producing pathogen specific
XX antigens of the invention.
XX SS Sequence 645 AA;
Query Match 11.1%; Score 89; DB 24; Length 645;
Best Local Similarity 21.4%; Pred. No. 4.2;
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;
QY 4 EVSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGYQF-----DGWEIS 54
Db 341 KMTDLQDTKYVYVESVENNESMMDTFVKH-----PIKTGMLNGKKYMMVMTTNDYWKDF 395
QY 55 GFEGKK-----DAGVINL-SKDTFI 74
Db 396 MVYEGQVRVTISKAKNTRTIFPVYEGKTLDAIVKVVHTIDYDGGVHVRIVDKAFT 455
QY 75 KPVFKKIEEKEENKPTFDV-----SKKKNQPNQVNHSQLNESHKEDLQ-----REEHSQ 125
Db 456 KANTDKSNKKEQDNSAKKEATPATPSKPTSPVEKESQKQDSQKDDNKQLPSVEKENDA 515
QY 126 KSDSTKQVTV--ATVLDKNNISSKSTNNPNK 154
Db 516 SSESCKGVTLATKPTKGEVESSTT--PTK 543
RESULT 29
ID ABB64828 standard; Protein; 281 AA.
XX AC ABB64828;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 21276.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
```



```
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-65860/75.
XX DR N-PSDB; ABL08931.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 21276; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 281 AA;
Query Match 11.1%; Score 88.5; DB 22; Length 281;
Best Local Similarity 22.1%; Pred. No. 1.5;
Matches 31; Conservative, 34; Mismatches 52; Indels 23; Gaps 5;
Qy 19 NGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKGADGYVI-----NLS 69
Db 137 NFEFTNMTTQDDLKNYI-SLLDK-----DNEGAIKSELGWMVIRALGRQPNESIA 188
Qy 70 KDTFTKPVFKKIEE-KKEEENKPTFDVSKKDNPOVNHSQLNE-----SHRKEDLQREEH 123
Db 189 KEFCNVLKRWHDTHKEELRDARFVFDKENNGYISITELRAVFMALGEXLEDELEEM 248
Qy 124 SKSDSTKDVATVLDKNNI 143
Db 249 IREYDLQDNHINFESNNM 268
RESULT 30
AAG85008
ID AAG85008 standard; Protein; 1141 AA.
XX AC AAG85008;
XX DT 11-SEP-2001 (first entry)
XX DE Shrimp white spot Bacilliform virus (WSBV) protein 99.
XX KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
XX KW antiviral agent; gene expression; antisense construct;;
XX KW transgenic viral resistant shrimp.
XX OS White spot syndrome virus.
XX OS WO200138351-A2.
XX PN 31-MAY-2001.
XX PD 08-NOV-2000; 2000WO-US28888.
XX PF 24-NOV-1999; 99CN-0124717.
XX PR (PENY-) PE CORP NY.
XX PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
XX PA (SINO-) SINOGENOMAX CO LTD.
```

---

```
PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
XX WPI; 2001-355877/37.
XX DR N-PSDB; AAG62788.
XX PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus
XX PT (WSBV), useful for producing viral polypeptides that can be used to
XX PT screen for agents that are useful for treating WSBV infection -
XX PS Claim 1; Figure 3; 626pp; English.
XX CC The invention provides the primary nucleotide sequence of the WSBV genome
XX CC (AAG62689), predicted transcript sequences (AAG62689-AAG62839) and
XX CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
XX CC (AAG62840-63160) suitable for use as primers or probes. The nucleic acid
XX CC molecules and proteins of the invention are useful for diagnosis and
XX CC monitoring viral infection, in screens for antiviral agents and for
XX CC monitoring viral gene expression or activity during a treatment regimen.
XX CC The nucleic acid molecules are also useful as antisense constructs to
XX CC control viral gene expression in infected cells and tissues and to create
XX CC transgenic viral resistant shrimp.
XX SQ Sequence 1141 AA;
Query Match 11.1%; Score 88.5; DB 22; Length 1141;
Best Local Similarity 23.3%; Pred. No. 11;
Matches 30; Conservative 31; Mismatches 43; Indels 25; Gaps 4;
Qy 25 STIVSEEDFILPVYKGELEKGYQFDGWEISGFEQ--KKDAGYVNLKDTFTKPVFKKIE 82
Db 621 SNVEESEE-----EEQWEEEBEVEEBEGSKEDDGDGDA-----PAQEEME 662
Qy 83 EKKEEENKPTFDVSKKDNPOVNHSQLNEHSHRKEDLQREHSHKSDSTKDVATVLDKNN 142
Db 663 EKKEEEOQ-----QQPEEENGENQEEEQQQQQQPEREENKADSDSDSDSSSSSSSS 717
Qy 143 ISSKSTNN 151
Db 718 SSSSSSSSS 726
RESULT 31
ABP73992
ID ABP73992 standard; Protein; 225 AA.
XX AC ABP73992;
XX DT 30-JAN-2003 (first entry)
XX DE Candida albicans essential protein SEQ ID NO 7829.
XX KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
XX KW signal transduction; DNA replication; cell division; growth;
XX KW proliferation; Candida albicans; fungicide; antifungal.
XX OS Candida albicans.
XX OS WO200253728-A2.
XX PN 11-JUL-2002.
XX PD 26-DEC-2001; 2001WO-US49486.
XX PF 29-DEC-2000; 2000US-259128P.
XX PR 20-FEB-2001; 2001US-0792024.
XX PR 22-AUG-2001; 2001US-314050P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX WPI; 2002-566694/60.
XX DR N-PSDB; ABZ32542.
```

XX Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele  
PT of a gene and placing other allele of the gene under conditional  
PT expression -  
XX  
PS Claim 44; SEQ ID NO 7829; 167pp + Sequence Listing; English.  
XX  
XX The invention relates to constructing (M1) a strain of diploid fungal  
XX cells in which both alleles of a gene are modified, comprising modifying  
XX one allele by insertion or replacement by a cassette having an  
XX expressible selectable marker and modifying other allele by  
XX recombination, of a promoter replacement fragment with a heterologous  
XX promoter, so that expression of the second allele is regulated by the  
XX promoter. (M1) is useful for constructing a strain of diploid fungal  
XX cells in which both alleles of a gene are modified. The diploid fungal  
XX cells having both alleles modified are useful for identifying a gene that  
XX is essential to the survival or growth of a fungus, a gene that  
XX contributes to the virulence and/or pathogenicity of a fungus, a gene  
XX that contributes to the resistance of a diploid fungus to an antifungal  
XX agent, an antifungal agent that inhibits the growth of a diploid fungus  
XX and for identifying a therapeutic agent for treatment of a mammalian  
XX disease. (M1) is useful for identifying a compound which modulates the  
XX activity of a gene product, preferably enzymatic activity, carbon  
XX compound catabolism, biosynthetic, transporter, transcriptional,  
XX translational, signal transduction, DNA replication and cell division  
XX activity. The method is useful for identifying a compound having the  
XX ability to inhibit growth or proliferation of C. albicans cells and for  
XX treating infection by C. albicans. The present sequence is that of an  
XX essential Candida albicans protein used in the method of the invention.  
XX Note: The sequence data for this patent is not represented in the printed  
XX specification but is based on sequence information supplied to Derwent by  
XX the European Patent Office.  
XX  
SQ Sequence 225 AA;  
Query Match 11.0%; Score 88; DB 23; Length 225;  
Best Local Similarity 28.0%; Pred. No. 1.2;  
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;  
QY 46 YQDGEWISGFEKGDAGVYVNLSDTKTFIKVPFKK----IEBKKEENKPTDFVSKKK-- 99  
DB 80 YDDDDDFGFESSNCAKXELNLSQAIEKWKQRDLIEREKLNSKKKEIEIKAKS 139  
QY 100 --DNPVNHSQLNESHKEDLOREEH--SQKSDSTKDVATVLDKNN 142  
DB 140 TIDDFYENTYNSKRDNHQKILSQEKFISKRDFFLK--RGTLDWRVN 184  
RESULT 32  
AAB94584  
ID AAB94584 standard; Protein; 258 AA.  
XX  
XX AAB94584;  
XX  
XX 26-JUN-2001 (first entry)  
XX  
XX Human protein sequence SEQ ID NO:15383.  
DE  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
XX Homo sapiens.  
XX  
XX EP1074617-A2.  
XX  
XX 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 15383; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 258 AA;  
Query Match 11.0%; Score 88; DB 22; Length 258;  
Best Local Similarity 26.3%; Pred. No. 1.5;  
Matches 31; Conservative 25; Mismatches 50; Indels 12; Gaps 4;  
QY 38 YKGELEKGYQDFGWEISGFEKGDAGY--VINLSKDTFIKVPFKKIEBKKEENKPTDFV 95  
DB 39 WEGEDEDEVDKNDWDDDDDEKKEAEVPEVKISEK--KKIAEKIKERQCKRQBEI 95  
QY 96 SKKKNPQ-----VNHSQLNESHKEDIQREHQSQKSDSTKDVATV--LDKNNTSSK 146  
DB 96 KKRLEEPKPKVLTPEQLADKLRLKQSSDLELAKETFGVNNVTYIGIDANPSSR 153  
RESULT 33  
AAM79318  
ID AAM79318 standard; Protein; 817 AA.  
XX  
XX AAM79318;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human protein SEQ ID NO 2964.  
DE  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
XX Homo sapiens.  
XX  
XX WO200157190-A2.  
XX

PD 09-AUG-2001.  
XX  
PF  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52451.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 215; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM8020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
SQ Sequence 817 AA;  
Query Match 11.0%; Score 88; DB 22; Length 817;  
Best Local Similarity 27.3%; Pred. No. 7.4; Mismatches 55; Indels 12; Gaps 4;  
Matches 33; Conservative 21;  
QY 40 GELEKGYQFDGWEISG--FEKKDAGYVINLSKDTFKPVPFKIEEKEEENKPTFDVSK 97  
Db 196 GQEKQESFKSWEASGKHQVSKPAVSLQKQDTSKLRSTLPEQKQEIISKSPSPSQ 255  
QY 98 -KKNPQVHSQLNESHKED-----LQREHHSQKSDSTKVTVATVLDKNNISSKT 148  
Db 256 WKQDTPKSKAGYVQEHKKQETPKLWVPVQLQKEQ-DPKKQTPKSWTPSMQSEQNTTKSWT 314  
QY 149 T 149  
Db 315 T 315  
RESULT 34  
AAM79319  
ID AAM79319 standard; Protein; 817 AA.  
XX  
AC AAM79319;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human protein SEQ ID NO 2965.  
DE  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW

KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorder; arthritis; inflammation.  
XX Homo sapiens.  
XX WO200157190-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52452.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 215; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM8020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
SQ Sequence 817 AA;  
Query Match 11.0%; Score 88; DB 22; Length 817;  
Best Local Similarity 27.3%; Pred. No. 7.4; Mismatches 55; Indels 12; Gaps 4;  
Matches 33; Conservative 21;  
QY 40 GELEKGYQFDGWEISG--FEKKDAGYVINLSKDTFKPVPFKIEEKEEENKPTFDVSK 97  
Db 196 GQEKQESFKSWEASGKHQVSKPAVSLQKQDTSKLRSTLPEQKQEIISKSPSPSQ 255  
QY 98 -KKNPQVHSQLNESHKED-----LQREHHSQKSDSTKVTVATVLDKNNISSKT 148  
Db 256 WKQDTPKSKAGYVQEHKKQETPKLWVPVQLQKEQ-DPKKQTPKSWTPSMQSEQNTTKSWT 314  
QY 149 T 149  
Db 315 T 315  
RESULT 35  
ABP73209  
ID ABP73209 standard; Protein; 635 AA.  
XX  
XX ABP73209;  
AC

XX 30-JAN-2003 (first entry)  
DE Candida albicans essential protein SEQ ID NO 7046.  
XX Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;  
KW signal transduction; DNA replication; cell division; growth;  
KW proliferation; Candida albicans; fungicide; antifungal.  
XX Candida albicans.  
XX WO200253728-A2.  
XX 11-JUL-2002.  
XX 26-DEC-2001; 2001WO-US49486.  
XX 29-DEC-2000; 2000US-259128P.  
PR 20-FEB-2001; 2001US-0792024.  
PR 22-AUG-2001; 2001US-314050P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KJ;  
PI WPI; 2002-566694/60.  
XX N-PSDB; AB231759.  
XX Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele  
PT of a gene and placing other allele of the gene under conditional  
PT expression  
XX  
XX Claim 44; SEQ ID NO 7046; 167pp + Sequence Listing; English.  
XX The invention relates to constructing (M1) a strain of diploid fungal  
XX cells in which both alleles of a gene are modified, comprising modifying  
XX one allele by insertion or replacement by a cassette having an  
XX expressible selectable marker and modifying other allele by  
XX recombination of a promoter replacement fragment with a heterologous  
XX promoter, so that expression of the second allele is regulated by the  
XX promoter. (M1) is useful for constructing a strain of diploid fungal  
XX cells in which both alleles of a gene are modified. The diploid fungal  
XX cells having both alleles modified are useful for identifying a gene that  
XX is essential to the survival or growth of a fungus, a gene that  
XX contributes to the virulence and/or pathogenicity of a fungus, a gene  
XX that contributes to the resistance and/or pathogenicity of a fungus, a gene  
XX agent, an antifungal agent that inhibits the growth of a diploid fungus  
XX and for identifying a therapeutic agent for treatment of a mammalian  
XX disease. (M1) is useful for identifying a compound which modulates the  
XX activity of a gene product, preferably enzymatic activity, carbon  
XX compound catabolism, biosynthetic, transporter, transcriptional,  
XX translational, signal transduction, DNA replication and cell division  
XX activity. The method is useful for identifying a compound having the  
XX ability to inhibit growth or proliferation of C. albicans cells and for  
XX treating infection by C. albicans. The present sequence is that of an  
XX essential Candida albicans protein used in the method of the invention.  
XX Note: The sequence data for this patent is not represented in the printed  
XX specification but is based on sequence information supplied to Derwent by  
XX the European Patent Office.  
XX Sequence 635 AA;  
XX  
Query Match  
Best Local Similarity 11.0%; Score 87.5; DB 23; Length 635;  
Matches 36; Conservative 38; Mismatches 68; Indels 21; Gaps 7;  
QY 4 EYSELKPH---RVTVTIQNGKM-----SSTIVSEDFILPVYKGELEKGYDFGW---51  
DB 70 KMKKFKPSPTWTPFPKQGIENLHWKSGKELIEQQEF---EEDGTPKPYFFKYNVQ 125  
QY 52 -EISGF--EGKDAGVIVLSK-DTPIKPVFKKIEEKEENKPFVDVSKKNDQNVNHS 107

Db 126 LEIPFVDETDYDL-YMIEIKYESKMKERARERKERKERKDLBEKKKQKQKQKS 184  
QY 108 QLNESHKEDLQREHHSQKSDSTKDVTVATVLDKNNISKSTTN 150  
DB 185 QONPQNKIKDKKQDTRNNTDKKDSQKSEKDKPTVEAKKETD 227  
RESULT 36  
ABBI12281  
ID ABB12281 standard; peptide; 2515 AA.  
XX ABB12281;  
XX 11-JAN-2002 (first entry)  
XX Human secreted protein homologue, SEQ ID NO:2651.  
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder; disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
KW antifungal; vulnery; antiulcer.  
XX Homo sapiens.  
XX WO200157188-A2.  
XX 09-AUG-2001.  
XX 05-FEB-2001; 2001WO-US03800.  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
XX (HVSE-) HVSEQ INC.  
XX Tang YT, Liu C, Drmanac RT;  
XX WPI; 2001-457740/49.  
XX N-PSDB; ABA09525.  
XX Human proteins and DNA encoding sequences useful for preventing,  
PT treating or ameliorating a medical condition in a mammalian subject  
PT e.g. arthritis and cancer  
XX  
XX Claim 20; Page 323-325; 1963pp; English.  
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
XX invention also relates to vectors and recombinant host cells comprising a  
XX nucleotide of the invention, methods of producing the novel polypeptides,  
XX antibodies against the polypeptides, methods of detecting the nucleotides  
XX or polypeptides in a sample, and methods of identifying compounds which  
XX bind to polypeptides of the invention. Although novel, many of the  
XX polypeptides of the invention have homology to known proteins, thereby  
XX giving an insight into their probable biological activities, and hence  
XX potential therapeutic applications. The polypeptides of the invention may  
XX have various activities, including cytokine, cell proliferation or cell  
XX differentiation activities; stem cell growth factor activity;  
XX haematopoiesis regulatory activity; tissue growth activity;  
XX immunomodulatory activity; activin- or inhibin-related activities;  
XX chemotactic or chemokinetic activities; haemostatic, thrombotic or  
XX thrombolytic activities; receptor or ligand activities; or may be  
XX involved in oncogenesis, cancer cell proliferation or metastasis.  
XX Depending on their biological activities, polypeptides and nucleotides of

the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

XX SQ Sequence 2515 AA;

Query Match 11.0%; Score 87.5; DB 22; Length 2515;

Best Local Similarity 24.2%; Pred. No. 40;

Matches 38; Conservative 29; Mismatches 55; Indels 35; Gaps 6;

QY 5 VSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGY 64

DB 2352 IAELOREPLLVNESLNVENSGFRTEEIHSESYNGKGEISSGRKONAEISGHSVEADP-- 2409

QY 65 VINLSKDTFIKPVFKKIEEKEENKPTFDVSKKK-----DNPQVNHSQLNESH 113

DB 2410 -----KEVEE--EERHMPK--RKRQHYLSSEDEPDNDPDLDSRI-ETA 2449

QY 114 RKEDLQREHQSQSDSTKDVATVLDKNISKSTTN 150

DB 2450 QRCQPEPETHATKEENSRDLEE--LPKTSSETNSTTS 2484

RESULT 37

AAM80268

ID AAM80268 standard; Protein; 2515 AA.

XX AC AAM80268;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3914.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;

XX KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

XX PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

DR N-PSDB; AAK53401.

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
XX useful in diagnosis and gene therapy -

PS Claim 20; Page 465-466; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.  
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
XX (AAM80020) are omitted as the relevant pages from the sequence listing  
XX were missing at the time of publication.

XX SQ Sequence 2515 AA;

Query Match 11.0%; Score 87.5; DB 22; Length 2515;

Best Local Similarity 24.2%; Pred. No. 40;

Matches 38; Conservative 29; Mismatches 55; Indels 35; Gaps 6;

QY 5 VSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGY 64

DB 2352 IAELOREPLLVNESLNVENSGFRTEEIHSESYNGKGEISSGRKONAEISGHSVEADP-- 2409

QY 65 VINLSKDTFIKPVFKKIEEKEENKPTFDVSKKK-----DNPQVNHSQLNESH 113

DB 2410 -----KEVEE--EERHMPK--RKRQHYLSSEDEPDNDPDLDSRI-ETA 2449

QY 114 RKEDLQREHQSQSDSTKDVATVLDKNISKSTTN 150

DB 2450 QRCQPEPETHATKEENSRDLEE--LPKTSSETNSTTS 2484

RESULT 38

AAM80269

ID AAM80269 standard; Protein; 2515 AA.

XX AC AAM80269;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3915.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;

XX KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

```
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR N-PSDB; AAK53402.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 466-467; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 2515 AA;
SQ
Query Match 11.0%; Score 87.5; DB 22; Length 2515;
Best Local Similarity 24.2%; Pred. No. 40;
Matches 38; Conservative 29; Mismatches 55; Indels 35; Gaps 6;
QY 5 VSELKPHRVTVTQNGKMSSTIVSEEDFILPVYKGELEKGYQFCGWEISGPEGKKDAGY 64
DB 2352 IAEQLQREPLVNESLVNENSGFTNEEIHSESYNGKEISSGRKDAEALSGHVSRAEP-- 2409
QY 65 VINLSKDTFKPVFKKIEEKEENKPTFDVSKK-----DNPQVNHSQLNESH 113
DB 2410 -----KEVEE--EERHMPK---RRKQHYLSSEDEPDNDPDLDSRI-ETA 2449
QY 114 RKEDLQREHSQKSDTKDVTATVLDKNNISKSTTN 150
DB 2450 QRQCPETEPHATKEENSRDLEE--LPKTSSTNSTTS 2484
RESULT 39
ABP75873
ID ABP75873 standard; Protein; 3021 AA.
XX
XX ABP75873;
XX
XX 10-FEB-2003 (first entry)
XX
XX Human secretory polypeptide SPTM SEQ ID NO 1057.
DE
XX Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;
KW asthma; Crohn's disease; neurological disorder; epilepsy; cancer;
KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;
KW multiple sclerosis; Parkinson's disease; cell proliferative disorder;
KW anti-inflammatory; immunosuppressive; neuroprotective; neurotropic;
KW neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;
KW antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;
KW secretory polynucleotide; protein.
XX
XX Homo sapiens.
OS
```

```
XX WO200283876-A2.
XX
XX 24-OCT-2002.
XX
XX 27-MAR-2002; 2002WO-US09921.
XX
XX 29-MAR-2001; 2001US-280067P.
PR 29-MAR-2001; 2001US-280068P.
PR 16-MAY-2001; 2001US-291280P.
PR 17-MAY-2001; 2001US-291829P.
PR 17-MAY-2001; 2001US-291849P.
PR 19-JUN-2001; 2001US-299428P.
PR 20-JUN-2001; 2001US-299776P.
PR 20-JUN-2001; 2001US-300001P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;
PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
XX WPI: 2003-075543/07.
DR N-PSDB; ABZ36315.
XX
XX New human secretory proteins and polynucleotides, useful for
PT diagnosing, treating or preventing autoimmune/inflammatory disorders
PT (e.g. AIDS), neurological disorders (e.g. Alzheimer's), or cell
PT proliferations or cancers -
XX
XX Claim 27; SEQ ID NO 1057; 458pp + Sequence Listing; English.
XX
XX The invention relates to a secretory polynucleotide (designated sptm)
CC comprising any of 567 polynucleotide sequences (ABZ35937-ABZ36403), a
CC naturally occurring polynucleotide sequence at least 90 % identical to
CC the polynucleotide sequence, a polynucleotide complementary to them or an
CC RNA equivalent of them. The polypeptide or polynucleotide are useful for
CC treating, preventing or diagnosing a disease or condition associated with
CC the expression of functional SPTM. These are particularly useful for
CC diagnosing, treating or preventing autoimmune/inflammatory disorders
CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's
CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,
CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,
CC schizophrenia or amnesia), or cell proliferative disorders (e.g.
CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,
CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,
CC breast, cervix or prostate). The present sequence is one of the SPTM
CC proteins of the invention (ABP75384-ABP75962).
CC
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3021 AA;
SQ
Query Match 11.0%; Score 87.5; DB 24; Length 3021;
Best Local Similarity 25.4%; Pred. No. 52;
Matches 43; Conservative 22; Mismatches 59; Indels 45; Gaps 7;
QY 5 VSELKPHRVTVTQNGKMSSTIVSEEDFILPVYKGELEKGYQFCGWEISGPEGKKDAGY 64
DB 2671 IAEQLQREPLVNESLVNENSGFTNEEIHSESYNGKEISSGRKDAEALSGHVSRAEP-- 2728
QY 65 VINLSKDTFKPVFKKIEEKEENKPTFDVSKK-----DNPQVNHSQLNESH 113
DB 2729 -----KEVEE--EERHMPK---RRKQHYLSSEDEPDNDPDLDSRIETAQ 2769
QY 114 R-----KEDLQR--EERHSQKSDTKDVTATVLDKNNISKSTT 149
DB 2770 RQCPETEPHATKEENSRDLEE--LPKTSSTNSTTSRVMEKDEYSSTT 2818
OS
```



Search completed: February 10, 2004, 10:53:55  
Job time : 37.1775 secs

```
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 10.9%; Score 87; DB 21; Length 313;
Best Local Similarity 25.4%; Pred. No. 2.5;
Matches 36; Conservative 26; Mismatches 46; Indels 34; Gaps 6;
QY 6 SELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISG-FEGKRDAGY 64
Db 55 SSEKPNR--KKIQKGKEIKSS-----PADG-KLSGKMKRKEKVG 91
QY 65 VINLSKDTFKPVFKIEEKKKEENKPTFDVSKKKNQPNVNHSQLNESHKEDLOREHS 124
Db 92 NYVDISEPILAEISTEKVKKEKGMNK-----TKKRAKEBITRSSV-----EDLKRESKF 141
QY 125 OKSDSTKDVTVTLVDKNNISSK 146
Db 142 KKSNNKKKMDMTSKKENKIEEE 163
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 29.0366 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848

Sequence: 1 TTVKEFLKNTGVESELKP.....ATVLDKNISSKSTNNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

- 1: sp archaea:\*
- 2: sp bacteria:\*
- 3: sp fungi:\*
- 4: sp human:\*
- 5: sp invertebrate:\*
- 6: sp mammal:\*
- 7: sp mhc:\*
- 8: sp organelle:\*
- 9: sp phage:\*
- 10: sp plant:\*
- 11: sp rodent:\*
- 12: sp virus:\*
- 13: sp vertebrate:\*
- 14: sp unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	848	100.0	2119	2 Q9AHT5	Q9AHT5 streptococc
2	848	100.0	2140	16 Q97RY6	Q97RY6 streptococc
3	845	99.6	2144	16 Q8DQP7	Q8DQP7 streptococc
4	844	99.5	2144	2 Q9S4M8	Q9S4M8 streptococc
5	119	14.0	775	16 Q8CPK8	Q8CPK8 straphylococ
6	112.5	13.3	361	5 Q9SPF15	Q9SPF15 plasmodium
7	112.5	13.3	379	5 Q9U6C4	Q9U6C4 plasmodium
8	111.5	13.1	346	5 Q9U0G0	Q9U0G0 plasmodium
9	111.5	13.1	354	5 Q25995	Q25995 plasmodium
10	111.5	13.1	354	5 Q81J55	Q81J55 plasmodium
11	111.5	13.1	379	5 Q25706	Q25706 plasmodium
12	111	13.1	1038	13 Q90784	Q90784 gallus gall
13	110.5	13.0	379	5 Q25705	Q25705 plasmodium
14	110	13.0	829	5 Q815F3	Q815F3 plasmodium
15	109.5	12.9	609	5 Q812K8	Q812K8 plasmodium
16	109	12.9	380	5 Q26019	Q26019 plasmodium

17	109	12.9	3008	5	Q81436	Q81436 plasmodium
18	108.5	12.8	600	5	Q77355	Q77355 plasmodium
19	107.5	12.7	470	10	Q9FUK9	Q9FUK9 arabidopsis
20	106	12.5	951	5	Q96229	Q96229 plasmodium
21	104.5	12.3	329	5	Q9NFV9	Q9NFV9 plasmodium
22	104	12.3	540	10	Q94C59	Q94C59 arabidopsis
23	104	12.3	540	10	Q9SA84	Q9SA84 arabidopsis
24	103.5	12.2	325	5	Q44016	Q44016 dictyosteli
25	103	12.1	2081	10	Q9LH98	Q9LH98 arabidopsis
26	102.5	12.1	238	5	Q81226	Q81226 plasmodium
27	101.5	12.0	312	16	Q9PPL5	Q9PPL5 campylobact
28	101.5	12.0	382	5	Q9V7J0	Q9V7J0 drosophila
29	101.5	12.0	556	5	Q9V7J9	Q9V7J9 drosophila
30	101.5	12.0	785	5	Q9GQ82	Q9GQ82 drosophila
31	101.5	12.0	948	5	Q9U4U6	Q9U4U6 plasmodium
32	101.5	12.0	1377	5	Q81556	Q81556 plasmodium
33	100.5	11.9	1373	5	Q81353	Q81353 plasmodium
34	100	11.8	375	4	Q14712	Q14712 homo sapien
35	100	11.8	622	16	Q98QA1	Q98QA1 mycoplasma
36	100	11.8	1130	5	Q81JZ4	Q81JZ4 plasmodium
37	100	11.8	3127	5	Q81DA0	Q81DA0 plasmodium
38	99.5	11.7	385	5	Q93424	Q93424 caenorhabdi
39	99.5	11.7	4524	5	Q813J9	Q813J9 plasmodium
40	99	11.7	211	5	P91488	P91488 caenorhabdi
41	99	11.7	219	16	Q9XOM6	Q9XOM6 thermotoga
42	99	11.7	1859	5	Q81C27	Q81C27 plasmodium
43	99	11.7	2162	5	Q81BH2	Q81BH2 plasmodium
44	99	11.7	2563	5	Q813A0	Q813A0 plasmodium
45	99	11.7	3026	5	Q81LS9	Q81LS9 plasmodium

#### ALIGNMENTS

#### RESULT 1

Q9AHT5 PRELIMINARY; PRT; 2119 AA.

AC Q9AHT5; DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Serine protease (Fragment).

GN PRTA.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=N4;

RX MEDLINE=21116976; PubMed=11179332;

RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,

RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,

RA Gayle A., Brewash Y.A., Walsh W., Warren P., Lathigra R., Hanson M.,

RA Langermann S., Johnson S., Koenig S.;

RT "Use of a Whole Genome Approach To Identify Vaccine Molecules

RT Affording Protection against Streptococcus pneumoniae Infection.";

RL Infect. Immun. 69:1593-1598(2001).

CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY

CC AN AMIDE BOND (BY SIMILARITY).

DR EMBL; AF291699; AAK19159.1; -.

DR HSSP; P00782; 2SPT.

DR MEROPS; S08.064; -.

DR InterPro; IPR001899; Gram\_pos\_anchor.

DR InterPro; IPR006192; LPXTG.

DR InterPro; IPR003137; PA.

DR InterPro; IPR000209; Peptidase\_S8.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00746; Gram\_pos\_anchor; 1.

DR Pfam; PF02225; PA; 1.

DR Pfam; PF00082; Peptidase\_S8; 2.

DR PRINTS; PR00723; SUBTILISIN.

DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.

```

DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Cell wall; Peptidoglycan-anchor; Protease.
FT NON_TER 1
SQ SEQUENCE 2119 AA; 238226 MW; 517997F6B960A6A CRC64;

Query Match 100.0%; Score 848; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 9,6e-53;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
Db 1922 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 1981

QY 61 WEISGFEKGDAGYVILNLSKDTFIKPVFKKIEKKEEENKPTFDVSKKDNQVNHSQLN 120
Db 1982 WEISGFEKGDAGYVILNLSKDTFIKPVFKKIEKKEEENKPTFDVSKKDNQVNHSQLN 2041

QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164
Db 2042 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2085

RESULT 2
Q97RY6 PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Serine protease, subtilase family.
GN SP0641.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIG84.
RX MEDLINE=21357209; PubMed=1146316;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umavam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506 (2001).
DR EMBL; AE007373; AAK74791.1; -.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRfam; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Protease; Complete proteome.

```

```

SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;

Query Match 100.0%; Score 848; DB 16; Length 2140;
Best Local Similarity 100.0%; Pred. No. 9.7e-53;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
Db 1943 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2002

QY 61 WEISGFEKGDAGYVILNLSKDTFIKPVFKKIEKKEEENKPTFDVSKKDNQVNHSQLN 120
Db 2003 WEISGFEKGDAGYVILNLSKDTFIKPVFKKIEKKEEENKPTFDVSKKDNQVNHSQLN 2062

QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164
Db 2063 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2106

RESULT 3
Q8DQP7 PRELIMINARY; PRT; 2144 AA.
AC Q8DQP7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN PrtA OR SP0561.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., BURGESS S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McMahon S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717 (2001).
DR EMBL; AE008434; AAK9365.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match 99.6%; Score 845; DB 16; Length 2144;
Best Local Similarity 99.4%; Pred. No. 1.6e-52;
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
Db 1947 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2006

QY 61 WEISGFEKGDAGYVILNLSKDTFIKPVFKKIEKKEEENKPTFDVSKKDNQVNHSQLN 120
Db 2007 WEISGFEKGDAGYVILNLSKDTFIKPVFKKIEKKEEENKPTFDVSKKDNQVNHSQLN 2066

QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164
Db 2067 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2110

RESULT 4
Q9S4M8 PRELIMINARY; PRT; 2144 AA.
AC Q9S4M8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
KW Protease; Complete proteome.

```

01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
Cell wall-associated serine proteinase PrtA precursor.  
PRTA.  
Streptococcus pneumoniae.  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN (1)  
SEQUENCE FROM N.A.  
RC STRAIN=3.B;  
RA Bethe G., ten Thoren E., Bongaerts R.J.M., Heinz H.-P., Zysk G.;  
RT "Cloning and sequencing of a novel surface protease of Streptococcus  
pneumoniae";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
AN AMIDE BOND (BY SIMILARITY).  
DR EMBL; AF127143; AA048399.1; -;  
DR HSSP; P00782; 2S8T.  
DR MEROPS; S08.064; -;  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Cell wall; Peptidoglycan-anchor; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 2144  
FT CDS 20 2144  
FT PROTEINASE  
FT CELL WALL-ASSOCIATED SERINE  
FT PRTA.  
SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
  
Query Match 99.5%; Score 844; DB 2; Length 2144;  
Best Local Similarity 98.8%; Pred. No. 1.9e-52;  
Matches 162; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 TTVKEFILNDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKYQFDG 60  
Dy 1947 TTVKEFILNDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKYQFDG 2006  
  
Qy 61 WEISGFEKKDAGYVINLSKDTFKIPVKFKIEEKEENKPTFDVSKKONPQVNHSQLN 120  
Dy 2007 WEISGFEKKDAGYVINLSKDTFKIPVKFKIEEKEENKPTFDVSKKONPQVNHSQLN 2066  
  
Qy 121 ESHRKEDLQREHKSQSDSTKVDTATVLDKNNISSKSTNNPNK 164  
Dy 2067 ESHRKEDLQREHKSQSDSTKVDTATVLDKNNISSKSTNNPNK 2110  
  
RESULT 5  
Q8CPK8 ID Q8CPK8 PRELIMINARY; PRT; 775 AA.  
AC Q8CPK8;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Penicillin-binding protein 1.  
GN SE0856.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,

RA Chen Z., Wen Y.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016746; AA004453.1; -;  
KW Complete proteome.  
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;  
  
Query Match 14.0%; Score 119; DB 16; Length 775;  
Best Local Similarity 27.0%; Pred. No. 1.1;  
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
  
Qy 10 KDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYK-----CELEKGYQFDG-- 61  
Dy 614 EDVNAQSLKP-----ITGNGKQIKQOSVKSGTKVLPHSKVLMGTDELTPW-DMTGWTK 668  
  
Qy 62 -ETSGFP-----GKKDAGYVIN--LSKDTFKIPVKFKIEEKEENKPTFDVS----K 107  
Dy 669 EDVLAFEDLTIKIVSTKNGFVTVNQSIKQIILK-----NKDKIEVLSAED 715  
  
Qy 108 KCONPQVNHSQLNESHKEDLQREHKSQSDSTKVDTATVLDKNNISSKSTNN 161  
Dy 716 TDDQEKTDSDSDNKSCKDADEHSNTSSSTKN-----DKSNADSKNDSDD 763  
  
RESULT 6  
Q95PI5 ID Q95PI5 PRELIMINARY; PRT; 361 AA.  
AC Q95PI5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Merzoite surface protein 3 (fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVO;  
RA Hisaeda H., Saul A., Long C.A., Miller L.H., Stowers A.W.;  
RT "Merzoite Surface Protein 3 and Protection Against Malaria in Aotus  
Monkeys";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY044180; AA094780.1; -;  
FT NON TER 1 1  
FT NON TER 361 361  
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;  
  
Query Match 13.3%; Score 112.5; DB 5; Length 361;  
Best Local Similarity 22.4%; Pred. No. 1.4;  
Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8;  
  
Qy 11 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKYQ----- 57  
Dy 147 ETGE-----RNSRNNFYTKTE-----YAKVEKDYERAKYAKNAQAV 187  
  
Qy 58 -----FD-----GWETSGF--EGKKDAG-----YVINLSKDTFKIPVKFKIEEKEEN 99  
Dy 188 LKAKEASSYDILGWEPGGVPEHKKENMLSHLYVSSKDKENISKENDVDLDE-KSEE 246  
  
Qy 100 KPTFDVSKKONPQVNHSQLNESHKEDLQREHKSQSDSTKVDTATVLDKNN 151  
Dy 247 EETEEELKEEKEEETEEISEDEEEEEEEEEEEEEEEEEEEEEEEEEEE 306  
  
Qy 152 --NISSKSTNN 161  
Dy 307 AQNLISKNNNN 318  
  
RESULT 7  
Q9U6C4 ID Q9U6C4 PRELIMINARY; PRT; 379 AA.  
AC Q9U6C4;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)



```
RA Perteau M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Bartell B.,
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014834; AAN35542.1; -.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 13.1%; Score 111.5; DB 5; Length 354;
Best Local Similarity 22.1%; Pred. No. 1.7;
Matches 41; Conservative 34; Mismatches 62; Indels 45; Gaps 7;

Qy 1 TTVEKILN-KDTGVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFD 59
Db 154 TKTEYAEKKNAYEKAKNAQKNAQAVLKAKEASS-----YDYL----- 194

Qy 60 GWEISGF--EGKQAG-----YVINLSKDTFTKPVFKKIEKKE-----EN 99
Db 195 GWFGGGVPEHKENMLSHLYVSSKDKENISKENDVDLDEKEEAEETBEELKEENEE 254

Qy 100 KPTFDVSKKNPQVNHQSHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTT 159
Db 255 ETEGSIKDEDEEEKEEENKKEQKEQSNENNNDQKQDMEA-----QNLISKQNN 309

Qy 160 NN 161
Db 310 NN 311

RESULT 11
Q25706 PRELIMINARY; PRT; 379 AA.
ID Q25706;
AC Q25706;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Polymorphic antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=C812;
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RL Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
DR Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; U08852; AAC47832.1; -.
SQ SEQUENCE 379 AA; 43302 MW; ABP9D54E1ED91A24 CRC64;

Query Match 13.1%; Score 111.5; DB 5; Length 379;
Best Local Similarity 23.6%; Pred. No. 1.8;
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;

Qy 19 KPHRVTVTIQNGKMSSTIVSEEDF-----ILPVYKGELEKGYQFD-GWEISGF- 66
Db 171 KPSRLNLSRKTKEAEQV--EKDYERAKNAYQKNAQAVLKAKEASSYDYLWFGGGV 228

Qy 67 -EGKQAG-----YVINLSKDTFTKPVFKKIEKKEENKPTFDVSKKKNPQVNHQSLN 120
Db 229 PEHKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEAEETBEELKEEKEETSEIS 287

Qy 121 -----BSHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNN 161
Db 288 EDEEEEEEKEEENKKEQKEQSNENNNDQKQDMEA-----QNLISKQNNN 336

RESULT 12
Q90784
```

```
ID Q90784 PRELIMINARY; PRT; 1038 AA.
AC Q90784;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Claustrian.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1];
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Cole G.J.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
RN [2];
RP SEQUENCE OF 1-451 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94157526; PubMed=7906711;
RA Burg M.A., Cole G.J.;
RT "Claustrian, an antiadhesive neural keratan sulfate proteoglycan, is
RT structurally related to MAP1B.";
RL J. Neurobiol. 25:1-22(1994).
DR EMBL; X67778; CAA47988.1; -.
SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;

Query Match 13.1%; Score 111; DB 13; Length 1038;
Best Local Similarity 23.7%; Pred. No. 5.9;
Matches 44; Conservative 28; Mismatches 58; Indels 56; Gaps 5;

Qy 3 VKFELNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKQDAGVYINLSKDTF 587
Db 528 VKQAKLQRTDSKESLAPAAKTTKQCKRNLKKHWSLQSLVQLEKPKQLESKEKTP 587

Qy 30 -----GKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKQDAGVYINLSKDTF 83
Db 588 VKKEKAVKPKETKTIKAEKDV-----TTKEQLGKSETSEKQASEKQDVKPKVTKES 639

Qy 84 IKPVFK-KIEKKEENKPTFDVSKKKNPQVNHQSHKEDLQRE-----EHSQKS 137
Db 640 VKGEKAKPEEKDEKPKKEVSKKEKPLI---KKEEKPKEDIKKEVKKEVKEK 696

Qy 138 DSTKDV 143
Db 697 EAKKEV 702

RESULT 13
Q25705 PRELIMINARY; PRT; 379 AA.
ID Q25705;
AC Q25705;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Polymorphic antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RL Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
DR Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; U08851; AAC47831.1; -.
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match 13.0%; Score 110.5; DB 5; Length 379;
Best Local Similarity 22.1%; Pred. No. 2.1;
Matches 43; Conservative 35; Mismatches 48; Indels 69; Gaps 9;
```

```

QY 11 DTGEVSELKPHRVTVVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQ----- 57
Db 167 ETGCE-----RNSRNNFYTTKKE-----YAGKVEKDYERAKNAYQKANOAV 207
QY 58 -----PD--GWEISGF--EGKKDAG-----YVINLSKDTFTKVPFKKEEKKEREN 99
Db 208 LKAKEASSYDILGWEFGGVPEHKKEENMLSHLYVSSKDKENISKENDVDVIDE-KEEA 266
QY 100 KPTFDVSKKNDPQVNHSLN-----BSHRKEDLQREHSQKSDSTKDVAT 146
Db 267 BETESEELEKNEETESISEDEEEEEEKKEKQKQSENNNDQKDMA- 325
QY 147 VLDKNNISSKSTNN 161
Db 326 ----QNLISKQNNN 336

RESULT 14
Q815F3 ID Q815F3 PRELIMINARY; PRT; 829 AA.
AC Q815F3;
DT 01-WAR-2003 (TremBLrel. 23, Created)
DT 01-WAR-2003 (TremBLrel. 23, Last sequence update)
DE 01-WAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFL1275C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Marten A., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014848; AAN36341.1; --
KW Hypothetical protein.
SQ SEQUENCE 829 AA; 98816 MW; EF2675E301B2CE93 CRC64;

Query Match 13.0%; Score 110; DB 5; Length 829;
Best Local Similarity 24.0%; Pred. No. 5.4;
Matches 43; Conservative 39; Mismatches 55; Indels 42; Gaps 9;

QY 2 TVKEFILNKDGEVSELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQFDGW 61
Db 491 TSNDLSILNTNKKI--IQP--LEYLLKNGD--KTLMTEDIVLDFPHFYMIKKYHLN-- 542
QY 62 EISGFGKDGAGVINLSKDTFTKVPFKKEEKNKPT-----FDVSK 107
Db 543 -----KKETLNFSLN-----FREIEKNRDKKKGTHNNKNDAEYWLKYIKK 587
QY 108 KKDN-PQVNHSLNESH-KREDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164
Db 588 KKKNFENNNTLNDNSNKKNNKLVEH---DNSLKQEQIILNDKNVIEHTKIYDNOKK 643

RESULT 15
Q812K8 ID Q812K8 PRELIMINARY; PRT; 609 AA.
AC Q812K8;
DT 01-WAR-2003 (TremBLrel. 23, Created)
DT 01-WAR-2003 (TremBLrel. 23, Last sequence update)

```

```

DT 01-WAR-2003 (TremBLrel. 23, Last annotation update)
DE Large cyclophilin-like protein.
GN PFI1490C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Gobie A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrall B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL929358; CAD51984.1; --
SQ SEQUENCE 609 AA; 72552 MW; 8CDF86E85FF9A021 CRC64;

Query Match 12.9%; Score 109.5; DB 5; Length 609;
Best Local Similarity 23.9%; Pred. No. 4.2;
Matches 42; Conservative 36; Mismatches 79; Indels 19; Gaps 5;

QY 5 EPIFNKDTGEVSELKPHRVTVTIQNGKEMSSIVSEEDFIL--PVYKGELEKGYQ 57
Db 114 KIILNRKTNVYSSEIYKL--FHNKNEFDPIDTKISKEDFVLQDPLNNKTSYVASH 170
QY 58 FDGWEISGFGKDGAGVINLSKDTFTKVPFKKEEKNKPTFDVSKKNDP----- 112
Db 171 LNDRDQEDNEKKKKNIFAPSIQDNGAIIKMEWEAKKERRKKEEKKKKYVIGIL 230
QY 113 QVNHSLNESHKEDLQREHSQKSDSTKDVATVLD-----KNNISSKSTNNPNK 164
Db 231 QDYNDSNESEHSNKKKMLKNKKDEYKNKIHSTNYDSSSENEIKICTNYSDNK 286

RESULT 16
Q26019 ID Q26019 PRELIMINARY; PRT; 380 AA.
AC Q26019;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DE 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE Polymorphic antigen precursor.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RX MEDLINE=95198774; PubMed=7891748;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
Plasmodium falciparum merozoites."
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
Plasmodium falciparum merozoite surface protein-3 (MSP-3).";

```

```

Query Match      12.9%; Score 109; DB 5; Length 3008;
Best Local Similarity 32.4%; Pred. No. 26;
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

Qy 74 YVINLSK-----DTPIKVFVKIEEKKEENKPTDVSKKDKDNPOVNHQS---LNESHKKE 126
      |||||
Db 2310 YDIELSKIEKFGASIGPVFTD-ENKKEENKN--EVNKKKEENKKEENKKEENKVK 2366
      |||||

```

RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RX MEDLINE=99087489; PubMed=9872454;  
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
RT physically assigned P1 and TAC clones.";  
RL DNA Res. 5:297-308(1998).  
DR EMBL; AB015468; BAB10694.1; --  
SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72E35AC54 CRC64;

Query Match 12.7%; Score 107.5; DB 10; Length 470;  
Best Local Similarity 20.1%; Pred. No. 4.4;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 19 KPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFE-----GKK 70  
DB 82 RENRVTDVTQNSGSK-----YVQDLARRIRYDE-EATGSQAQRIDHPNOK 129  
QY 71 DAGYVNLKDTFIKVPFKKIEBKEEENKPTFDVSKKDN----- 111  
DB 130 NVGITEKAPENGPIEBTSDNDKRNQKNNKFTAAKSSENAVSRVSFGADHKRAEVMGK 189  
QY 112 PQVNSHQLNE-----SHRKDLQREHSQKSDSTKDVTAIVLDKNNISSKSTNNPNK 164  
DB 190 PMENRDQVRQTESAEKSHRKNVTSEKPRDQGVKKTEAKDKDRNKEKKEETESINK 248

RESULT 20  
O96229 ID O96229 PRELIMINARY; PRT; 951 AA.  
AC O96229;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN FFB0680W.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=99021743; PubMed=9804551;  
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,  
RA Koonin E.V., Shalton S., Mason T., Yu K., Fujii C., Pederson J.,  
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,  
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.";  
RL Science 282:1126-1132(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=22255705; PubMed=1236864;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Paulsen K.E., Bowman S., Paulsen I.T., James K.,  
RA Eissen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shalton S.J., Suh B., Peterson J., Angiuoli S.,  
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrall B.;  
RT "Genome sequence of the human malaria parasite Plasmodium falciparum.";  
RL Nature 419:498-511(2002).  
RL EMBL; AB001410; AAC71925.2; --  
KW Hypothetical protein.  
SQ SEQUENCE 951 AA; 112486 MW; AC6D889358A84F4F CRC64;

Query Match 12.5%; Score 106; DB 5; Length 951;  
Best Local Similarity 24.3%; Pred. No. 12;  
Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;

QY 10 KDTGVSFLKPHRVT-VTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEI--SGF 66  
DB 128 EEKNKINKSDDLHRQNELNLQSGK-----NEQDI-----NKNEKGKQ----DISNSNA 170  
QY 67 EGKQDAGVYVNLKDTFIKVPFKKIEBKEE-----ENKPTFD-----VSKKKNP 112  
DB 171 ENKKD-----VKEGVKELEEKKEBKISDDHKVBEENKSDDKHKVEENKSDDH 218  
QY 113 QVNSHQLNESHKEDLQREH-----SQQSDSTKDVTAIVLDKNNISSKSTNNPNK 164  
DB 219 KVEENKSDDKHKIEEVKVEEHEDEE-----DKKESKSNKKNKDNK 262

RESULT 21  
O9NFV9 ID O9NFV9 PRELIMINARY; PRT; 329 AA.  
AC O9NFV9;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
GN MSP3.  
OS Plasmodium falciparum (isolate 7G8).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=57266;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=7G8;  
RX MEDLINE=20416497; PubMed=10960178;  
RA Okenu D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in Plasmodium reichenowi and Plasmodium falciparum.";  
RL Mol. Biochem. Parasitol. 109:185-188(2000).  
RL EMBL; AJ252287; CAB85901.1; --  
FT NON\_TER 1  
FT NON\_TER 329  
SQ SEQUENCE 329 AA; 36916 MW; C5B045DB5E21A159 CRC64;

Query Match 12.3%; Score 104.5; DB 5; Length 329;  
Best Local Similarity 24.3%; Pred. No. 4.9;  
Matches 43; Conservative 30; Mismatches 63; Indels 41; Gaps 8;

QY 1 TTVKEFILN-KDTGVSFLKPHRVT-VTIQNGKMSSTIVSEDFILPVYKGELEKGYQFD 59  
DB 144 TKTKEYAEKAKNAYEAKNAQKAVLKAKEASS-----YDYL----- 184  
QY 60 GWEISGF--EGKQDAG-----YVNLKDTFIKVPFKKIEBKEEENKPTFDVSKKKNP 112  
DB 185 GWEFGGVPEHKKEENMLSHLYVSKKENISKENDVDLDE--KEEAEETEEELKEKNE 243  
QY 113 QVNSHQLNESHKEDLQREH-----SQQSDSTKDVTAIVLDKNNISSKSTNN 161  
DB 244 ETESEISEDEEEEEEKKEEKEQAKESNNNDKQDMEAE-----QNLISKQNNN 295

RESULT 22  
O9AC59 ID O9AC59 PRELIMINARY; PRT; 540 AA.  
AC O9AC59;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Putative phosphatidyl-inositol-transfer protein.  
GN T518.14.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;



```

[1]
RN SEQUENCE FROM N.A.
RP Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
RA Bower L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.,
RT "Full Length cDNA of gene TS18.14 (GI:4587525).";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY035162; AAKS9666.1; -.
DR InterPro; IPR001251; CRAL_TRIO.
DR Pfam; PF00650; CRAL_TRIO; 1.
DR PRINTS; PR00180; CRETINALDHP.
DR SMART; SM00516; SEC14; 1.
DR PROSITE; PS0191; CRAL_TRIO; 1.
SQ SEQUENCE 540 AA; 61166 MW; 0C2592EB18ACFAC6 CRC64;
Query Match 12.3%; Score 104; DB 10; Length 540;
Best Local Similarity 25.3%; Pred. No. 9.2;
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;
QY 1 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60
DB 27 TTVKA-VVEETKVEDESKP-----EGVEKSAFKEESDFPADLKESEKK----- 70
QY 61 WEISGFEKKDAGYVINLSKDTFIKVPFK-----IEEKEEENKPTFDVSKKKDNQVNH 117
DB 71 -ALSDLKSLKEAIVDN-----TLTKTKKSSPMKEKEEVVYKPAEVEKKKE--EAAEE 123
QY 118 OLNSHREKEDLQREHSOKSDSTKDVATVLDKNNISS 155
DB 124 KVEEKKSEAVVTEAPKAETVEAVVTEIIPKEEVT 161
Query Match 12.3%; Score 104; DB 10; Length 540;
Best Local Similarity 25.3%; Pred. No. 9.2;
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;
QY 1 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60
DB 27 TTVKA-VVEETKVEDESKP-----EGVEKSAFKEESDFPADLKESEKK----- 70
QY 61 WEISGFEKKDAGYVINLSKDTFIKVPFK-----IEEKEEENKPTFDVSKKKDNQVNH 117
DB 71 -ALSDLKSLKEAIVDN-----TLTKTKKSSPMKEKEEVVYKPAEVEKKKE--EAAEE 123
QY 118 OLNSHREKEDLQREHSOKSDSTKDVATVLDKNNISS 155
DB 124 KVEEKKSEAVVTEAPKAETVEAVVTEIIPKEEVT 161
RESULT 23
Q9SA84
ID Q9SA84 PRELIMINARY; PRT; 540 AA.
AC Q9SA84;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE T518.14 protein (Hypothetical protein).
GN T518.14 OR ATIG30690.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=sv. Columbia;
RA Vysotskaia V.S.; Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Kremenetskaia I., Luros J., Ngan I., Gonzalez A., Altafi H.,
RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federpiel N.A., Theologis A.,
RT "Arabidopsis thaliana chromosome 1 BAC T518 sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Huan V.W., Lee J.M.,
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007060; AAD25756.1; -.
DR EMBL; BT000959; AAN41359.1; -.

```

```

DR InterPro; IPR001251; CRAL_TRIO.
DR InterPro; IPR001071; RetBind/tocTrans.
DR Pfam; PF00650; CRAL_TRIO; 1.
DR Pfam; PF03765; CRAL_TRIO N. 1.
DR PRINTS; PR00180; CRETINALDHP.
DR SMART; SM00516; SEC14; 1.
DR PROSITE; PS0191; CRAL_TRIO; 1.
KW Hypothetical protein.
SQ SEQUENCE 540 AA; 61189 MW; 0C2590D518ACFB58 CRC64;
Query Match 12.3%; Score 104; DB 10; Length 540;
Best Local Similarity 25.3%; Pred. No. 9.2;
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;
QY 1 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60
DB 27 TTVKA-VVEETKVEDESKP-----EGVEKSAFKEESDFPADLKESEKK----- 70
QY 61 WEISGFEKKDAGYVINLSKDTFIKVPFK-----IEEKEEENKPTFDVSKKKDNQVNH 117
DB 71 -ALSDLKSLKEAIVDN-----TLTKTKKSSPMKEKEEVVYKPAEVEKKKE--EAAEE 123
QY 118 OLNSHREKEDLQREHSOKSDSTKDVATVLDKNNISS 155
DB 124 KVEEKKSEAVVTEAPKAETVEAVVTEIIPKEEVT 161
RESULT 24
O44016
ID O44016 PRELIMINARY; PRT; 325 AA.
AC O44016;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE G5 ORF.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NC4;
RX MEDLINE=98198836; PubMed=9539429;
RA Rieben W.K. Jr., Gonzales C.M., Gonzales S.T., Pilkington K.J.,
RA Kiyosawa H., Hughes J.E., Welker D.L.;
RT "Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to
the Ddp1 and Ddp2 plasmid families.";
RL Genetics 148:1117-1125(1998).
DR EMBL; U00796; AAC18634.1; -.
SQ SEQUENCE 325 AA; 38448 MW; 69A43D0C632058A6 CRC64;
Query Match 12.2%; Score 103.5; DB 5; Length 325;
Best Local Similarity 23.8%; Pred. No. 5.7;
Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;
QY 21 HRVTVTIQNGKEMSTIVSEEDFILPVYK-GEL--EKGYQFDGWEISGFEK----- 69
DB 57 HRTTITSIKN--RFVSKXIGDEKELFRISKNGELIVLNELEFDNFHIK--EGHLRKSKWF 112
QY 70 ---KDAGV-----INLSKDTIKPV-----FKK----- 90
DB 113 NHIKDSGYATNIEIFLESTCLCKTAQTKNSYKKNRIINKLPEEEEEEEEEEE 172
QY 91 --IEEKEEENKPTFDVSKKKDNQVNHSQLNSHREK-----DLQREHSOKSDSTK 141
DB 173 EEESEEEVEKPTISEEEEEETPAVSEEEEEEEEEEEETPAVSEEEEEEEQEDKEK 232
QY 142 D 142
DB 233 D 233
RESULT 25
Q9LH98

```

```
ID Q9LH98 PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T13N8.
DS Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1] SEQUENCE 2081 AA; 232851 MW; D3603E1F85EFFF29 CRC64;
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones";
RL DNA Res. 7:217-221(2000).
DR EMBL; AF002057; BAB03174.1; -.
SQ SEQUENCE 2081 AA; 232851 MW; D3603E1F85EFFF29 CRC64;

Query Match 12.1%; Score 103; DB 10; Length 2081;
Best Local Similarity 26.0%; Pred. No. 47;
Matches 46; Conservative 29; Mismatches 72; Indels 30; Gaps 7;

QY 7 ILNKDTGVESELK---PHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEI 63
Db IVERNGKEDSIKESGDKTVEINGEELSTEEGSKD-----GKIEEGK--EGKEN 1688
QY 64 SFEGEKD-----AGYVNLKSDTFIKVPFKIEEKEENKPTFDVSK-KKD 110
Db STKEGSKDKIKIEGMEKENSTKESKDKINEIHGDKKEATMEGSKGGTSTGKDSKD 1748
QY 111 NPQVNHSQLNESHKEDLQ---REBHSQKSDTKDVTATVLDKNNISSKSTNNPN 163
Db SKSVEINGVDDSLKDDSKNGDINEINCKEDSVKDNVTIEIQGNDNSLNTSISSEPN 1805

RESULT 26
Q81226 PRELIMINARY; PRT; 238 AA.
AC Q81226;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN Pfl0765W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1] SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Ackin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
```

```
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929356; CAD51839.1; -.
KW Hypothetical protein.
SQ SEQUENCE 238 AA; 28481 MW; DA175A2B5A109B49 CRC64;

Query Match 12.1%; Score 102.5; DB 5; Length 238;
Best Local Similarity 28.0%; Pred. No. 4.8;
Matches 49; Conservative 26; Mismatches 65; Indels 35; Gaps 10;

QY 10 KDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISG---F 66
Db KEKEXTQETKHD-----ENKSNVNVNSENKPTFDVSKKKDN--POV--N 91
QY 67 EGKDA-----GYVINLSKDTFIKVPFKIEEKEENKPTFDVSKKKDN--POV--N 115
Db ESKKATSRKKIKLNFKLKNNLIEKVIK--KAPLEAKNTVFPSLKKNYTVPHISNN 148
QY 116 H--SQLNESHKED-----LQREHSQKSDTKDVTATVLDKNNISSKSTNNPN 163
Db HNISQNDHNKKKQDKNIKLLKLEKSNKDDNN-----NINNKNNNQKCDTKNLN 198

RESULT 27
Q9PPL5 PRELIMINARY; PRT; 312 AA.
AC Q9PPL5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Putative membrane protein.
GN CJO692C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1] SEQUENCE FROM N.A.
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139076; CAB72966.1; -.
KW Complete proteome.
SQ SEQUENCE 312 AA; 37221 MW; 0004FA7836A741E8 CRC64;

Query Match 12.0%; Score 101.5; DB 16; Length 312;
Best Local Similarity 25.0%; Pred. No. 7.6;
Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;

QY 5 EFLINKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYK-----GELE 53
Db DIKERQNIQKWLKFSR---AINQKPLDDDLRDEISSDDILRRFKKTKTNKFLLELD 61
QY 54 KYQFDGWEISGFEKGKAGYVINI---SKDTFIKVPFKIEEKEENKPT---FDVSK 107
Db EYESKHTKSKSYLKD---LINVLEKQSLAKKIFSKMKERKEENKTKKQFLFSR 118
QY 108 KDNP---QYNHSQLNESHKEDLQREHSQKSDTKDV---TATVLDKNNISSK--STT 159
Db KKAENIKNIQTQITQTSNQATTTQKKELTNSIEIKTETKIQKPLIEKLDVK 178
QY 160 NNPEN 163
Db 179 NQPN 182
```

RESULT 28

Q9V7J0 PRELIMINARY; PRT; 382 AA.

AC Q9V7J0; Q9G081.

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)

DE CG8421 protein (Aspartyl beta-hydroxylase variant 2).

GN ASPH OR CG8421 OR CG18658.

OS *Drosophila melanogaster* (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

ON NCBI\_TaxID=7227;

OX [1]

RN SEQUENCE FROM N.A.

RP STRAIN=Berkely;

RC MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Anril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.J., McCleod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "the genome sequence of *Drosophila melanogaster*;"

RL Science 287:2185-2195 (2000).

RN [2]

RN SEQUENCE FROM N.A.

RP Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Waesman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.O.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*,"  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banjon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferriera S., Frick E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svitskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
RT "Sequencing of *Drosophila melanogaster* genome,"  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Mixra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.I., Brydasdale R., Berman B., Carlson J.W., Celniker S.E.,  
RA Clump M., Drgasdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
RT "Annotation of *Drosophila melanogaster* genome,"  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AF5003808; AAF5064.2; -

DR FlyBase: FBgn0034075; Asph.  
SQ SEQUENCE 556 AA; 63144 MW; B420980CDB6C357A CRC64;  
Query Match 12.0%; Score 101.5; DB 5; Length 556;  
Best Local Similarity 24.5%; Pred. No. 14;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
QY 41 EHFILPVVKGLEKGYQFDGW-----HISGFEKKDAGYVI-----NLKDTFTK 85  
DB 78 ELDLPLSESFSK--VFDGWVDEHRDHDGHDVQPSGEALDDHDDHDDHDEDEE 135  
QY 86 PVFKIEEKKEENKPT-----PVSKKKNDPQVNHSHKQLSHKRELQREHSKSDS 139  
DB 136 PITELELELEEEETFEDEPAADVEEYDEDEENNA--GENITAEDAEEEDND 193  
QY 140 TKDVTATVLDKNNISKST 158  
DB 194 EGTVEATVEATTEATTEAT 212  
RESULT 30  
Q9GQ82 PRELIMINARY; PRT; 785 AA.  
AC Q9GQ82; DT 01-MAR-2001 (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA).  
GN ASPH OR CG8421 OR CG18658.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
ON NCBI Taxid=7227;  
RX [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20564328; PubMed=10956665;  
RA Dinchuk J.B., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
RA Friedman P.A.;  
RT "Aspartyl beta-hydroxylase (Asph) and an evolutionarily conserved  
RT Isoform of Asph Missing the Catalytic Domain Share Exons with  
RT Junction.";  
RL J. Biol. Chem. 275:39543-39554 (2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Waesman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.O.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*,"  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banjon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferriera S., Frick E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svitskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
RT "Sequencing of *Drosophila melanogaster* genome,"  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Mixra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.I., Brydasdale R., Berman B., Carlson J.W., Celniker S.E.,  
RA Clump M., Drgasdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
RT "Annotation of *Drosophila melanogaster* genome,"  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AF5003808; AAF5064.2; -

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Pallazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
[3]  
RP SEQUENCE FROM N.A.  
RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banton J., Beeson K.Y., Buesam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferreira S., Frise E., Gallie R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragov V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Mira S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celnik S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[6]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF289493; AAG40806.1; -  
DR EMBL; AE003808; AAM70947.1; -  
DR FlyBase; FBgn0034075; Asph.  
DR InterPro; IPR001440; TPR.  
DR InterPro; IPR006025; Zn MP1ptdase.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
SQ SEQUENCE 785 AA; 89843 MW; 30A8DFCD6836F7F1 CRC64;  
Query Match 12.0%; Score 101.5; DB 5; Length 785;  
Best Local Similarity 24.5%; Pred. No. 21;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
QY 41 EDFILPVYKGEKCYQFDGM-----EISGEGKDKAGYVI-----NLSKDTFIK 85  
DB 78 EDLDTPLSESRFSK--VFQGVDEHRDHDGHVQVGEALDDHDEHDDHDDHDEDEE 135  
QY 86 PVFKIEKKEENKPT-----FVSKKKKPNQVHSQLNESHKEDLQREHSQKSDS 139  
DB 136 PLTELEEELEEEPTDEDEPAADDEYEDEDEENNA--GENITAEADAESEEDND 193  
QY 140 TKDVTATVLDKNNISKST 158  
DB 194 EGTVEATVEATTEAT 212

RESULT 31  
Q9U4U6 PRELIMINARY; PRT; 948 AA.  
AC Q9U4U6;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Ornithine decarboxylase.  
GN ODC  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FCUP1/RS;  
RA Birkholtz L., Joubert F., Neitz A.W.H., Louw A.I.;  
RT "Molecular characterisation of Plasmodium falciparum ornithine  
RL decarboxylase cDNA obtained by RACE.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF139900; AAF14518.1; -  
DR HSSP; P07805; 1F3T.  
DR InterPro; IPR000183; Decarboxylase2.  
DR Pfam; PF02784; Orn Arg dec N; 1.  
DR Pfam; PF02078; Orn DAP Arg dec; 1.  
DR PRINTS; PR01179; ODADCRXLASE.  
SQ SEQUENCE 948 AA; 110350 MW; 43F103DB83F12835 CRC64;  
Query Match 12.0%; Score 101.5; DB 5; Length 948;  
Best Local Similarity 21.8%; Pred. No. 26;  
Matches 51; Conservative 31; Mismatches 63; Indels 89; Gaps 11;  
QY 6 FILAKDTEVSEL-----KPRVTVTIQTNGKMSSTIVSEEDFILPVY-----KG 50  
DB 559 YIIILGGYPEELSYDNAKGDKHYCTLSQEIKKOIQKFLNEETFKTKGYYSFKI 618  
QY 51 ELERGYQFDGWEISGFEKGD-----AGYVINLSKDTFIKPVFKIEKKEENKPT 102  
DB 619 SLAINMSIDHY-----FSHMKNLRAVICPGRYVAAASSTLAVKIIGK-----RPT 665  
QY 103 F-----DVSKKNPNQVHSQLNESHKED-----LQREHSQK----- 136  
DB 666 FQGIIMLKDLKAHYDPLNFAQOENKQDEPKINHNNNDNNNDNNNNNNNNKQGG 725  
QY 137 -----SDST-----KDYATATVLDK--NNIS-SKSTTNNPN 163  
DB 726 GNIMNDLIITSTNDSTNKNKNDHSSQVIONVSVCTIRKEGDKINHTHTINPN 779  
RESULT 32  
Q81556 PRELIMINARY; PRT; 1377 AA.  
AC Q81556;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Eukaryotic translation initiation factor 3 subunit 10, putative.  
GN PFI0625C.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=22255705; PubMed=12368864;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
RA Perteira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

```
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RL Nature 419:498-511(2002).
RL EMBL; AF014846; AAN36214.1; -.
KW Initiation factor.
SQ SEQUENCE 1377 AA; 166059 MW; 587CF2E3F2C8FB99 CRC64;

Query Match 12.0%; Score 101.5; DB 5; Length 1377;
Best Local Similarity 22.3%; Pred. No. 39;
Matches 43; Conservative 37; Mismatches 62; Indels 51; Gaps 9;

QY 3 VKFELNKDTGSELSKPHRVTVTTONGKE--MSSTIVSEEDFILPVYKGELEKGYQF-- 58
DB 619 IKDSLINMYNDLQNGLKMTNIAIENEVENITMSTLTYYEEEF--AYLDKSNKNINVT 675
QY 59 ----DGEWISGPEGKDDAGVINLSKDTFKPVFKIEBK-----KEEENKPTFDVSK 107
DB 676 ATLEDIIEVEFEKP-----IENEKLLKIIYKIDDEHHKIQLLSEHNKKRKLK 728
QY 108 KKDNPQVNHSQLN-----ESHKEDLQR-----BEHSQKSDSTKDVAT 146
DB 729 KQ--KELEQAQLKMKELKLEKEKELARKEKRIKEEKKHKKTEAAEQMLKE 786
QY 147 VLDKNNISSKSTT 159
DB 787 I---KKLCSTNTT 796

RESULT 33
Q81353 PRELIMINARY; PRT; 1373 AA.
AC Q81353;
DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Helicase with Zn-finger motif, putative.
GN PF10480W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moulé S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultston J.E., Craig A., Newbold C., Barrell B.G.
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL929356; CAD51782.1; -.
SQ SEQUENCE 1373 AA; 160372 MW; 059FB4E99EEF4702 CRC64;

Query Match 11.9%; Score 100.5; DB 5; Length 1373;
Best Local Similarity 21.7%; Pred. No. 45;
Matches 40; Conservative 33; Mismatches 54; Indels 57; Gaps 9;

QY 6 FILNKDTGTVSE----LKPHRVTVTTIONGKMSSTIVSEEDFILPVYKGELEKGYQFGW 61
DB 418 YVNDGRGVWEESIIMPLPHVHQLL-----LSATV-----PNY-----LEFADW 457
QY 62 -----EISGPEGKK-----DAGVINLSKDTFKPVFK-----KIEKKKEE 98
```

```
Db 458 VGFTQKEVISISTKKRPVPLHYVYVDSVYLVMDENKKNFYSSAFKELYVYKIREQEAN 517
QY 99 NKPTFDVSKKDNPOVNHSQLNESH-RKEDLQREHSHQSKSDSTKDVATVLDKNNISSKS 157
Db 518 NKNTKQITSGSNTSSNLKNNNYYSKKNYLTNTNNKENDNT-----QNNINNNN 568
QY 158 TTNN 161
Db 569 NNNN 572

RESULT 34
Q14712 PRELIMINARY; PRT; 375 AA.
AC Q14712;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Cell cycle progression restoration 8 protein.
GN CPR8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98043401; PubMed=9383053;
RA Edwards M.C., Liegeois N., Horecka J., Depinho R.A., Sprague G.F. Jr.,
RA Tyers M., Elledge S.J.
RT "Human CPR (cell cycle progression restoration) genes impart a Far-
phenotype on yeast cells."
RL Genetics 147:1063-1076(1997).
DR EMBL; AF011794; AAB69314.1; -.
DR InterPro; IPR004238; LEA.
DR Pfam; PF02987; LEA; 1.
SQ SEQUENCE 375 AA; 44320 MW; 0C261BF68AAEF7AD CRC64;

Query Match 11.8%; Score 100; DB 4; Length 375;
Best Local Similarity 28.0%; Pred. No. 12;
Matches 51; Conservative 16; Mismatches 67; Indels 48; Gaps 9;

QY 13 GEVSELKPHRVTVTTIONGKMSSTIVSE--EDFILPVYKGELEKGYQF-----DGWE--- 62
Db 18 GELQOLSGSQL-----HGKSDSPNVYTEKKAIRLRRLTELRKLTTEQQRSDLWRLY 72
QY 63 -----ISGPEGKKDAG---YVINLSKDTFKPVFKIEKKEEENKPTFDVSKKKD 110
Db 73 VEAKDQNGKQTDGKKGGRGSHRVKNKSKGTFLGSV-----KETFDAMKNST 120
QY 111 NPQVNHSQLNESHKEDLQREHSHQSKSDST-----KDVATVLD-KNNISSKSTTNP 162
Db 121 KEFVRHHEKIKQAKEDV-KENLKKFSDSVKSTPRHFDTTKNIFDEKGNKFNATKEA 179
QY 163 NK 164
Db 180 EK 181

RESULT 35
Q98QAL PRELIMINARY; PRT; 622 AA.
AC Q98QAL;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE LIPOPROTEIN.
GN MYP_4650.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RESULT 37
Q8IDA0 PRELIMINARY; PRT; 3127 AA.
ID Q8IDA0
AC Q8IDA0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Guanylyl cyclase [EC 4.6.1.2].
GN GC-BETA.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA Barriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844509; CAD52725.1; -.
KW Lyase.
SQ SEQUENCE 3127 AA; 370597 MW; F0375C72B9007560 CRC64;
Query Match 11.8%; Score 100; DB 5; Length 3127;
Best Local Similarity 24.2%; Pred. No. 1.2e+02;
Matches 52; Conservative 34; Mismatches 71; Indels 58; Gaps 9;
QY 4 KEFILNKDGEVSELKPHRVTVTIQNG-----KEMSSTVSEDFILPVYKGELEKGYQ 57
DB 1969 KKF--KRNTSYVLESPLHLIGDIVDNNIRKKKKKEIKTVSDMTSPVNIKEYNYNEQ 2026
QY 58 FDGWEISG---PEGKKDAGYVNLKSDTKPIKPVFKIEKKKEENK----- 100
DB 2027 ERKKEIVGNLSYDKTKKIPFFIKFTKEGRKK--KKIEKKEKKEKNNNNFLYNDYSS 2084
QY 101 ---PTF-----DVSKKNQPNVHSQL-----NESHK---EDLQR 130
DB 2085 YSFPKYGDNNENFVIKVIERKDFQCKFDHPNFNFKFLHYNPMKNKKNKKNKNVRR 2144
QY 131 EEHSQKSDSTKO-VTATVLDKNNISSKSTTNNPNK 164
DB 2145 NEYPNTYSSKDGVSYNFLSDSLFSSDNEYSSDNE 2179
RESULT 38
Q93424 PRELIMINARY; PRT; 385 AA.
ID Q93424
AC Q93424; P90801;
DT 01-JAN-1999 (TrEMBLrel. 09, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical glycine-rich 37.0 kDa protein E02A10.2 in chromosome V precursor.
GN E02A10.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;
RA STRAIN=BRISTOL N2;
RA Thomas K.; (OCT-1996) to the EMBL/GenBank/DBJ databases.
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RP STRAIN=BRISTOL N2;
RA Jones S.J.M.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z81053; CAB02877.1; -.
DR HSSP; P10968; 2CWG.
DR WormPep; E02A10.2; CE09116.
DR InterPro; IPR002952; Eggshell.

```

Search completed: February 10, 2004, 10:57:09  
Job time : 31.0366 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 11.8417 Seconds  
(without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Sequence: 1 TTVEKFLINKDTGEVSELKP.....ATVLDKNNISSKSTNNPNK 164

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	2140	2 F95074	serine proteinase,
2	845	99.6	2144	2 A97942	metalloproteinase
3	111	13.1	1038	2 JCS497	claustrin - chick
4	110	13.0	558	2 T18467	hypothetical prote
5	106	12.5	665	2 B71609	hypothetical prote
6	104	12.3	540	2 D86432	hypothetical prote
7	103.5	12.2	325	2 T18283	hypothetical prote
8	101.5	12.0	312	2 G81339	probable membrane
9	100	11.8	622	2 A90570	lipoprotein (impor
10	99.5	11.7	385	2 T20410	hypothetical prote
11	99	11.7	211	2 T25911	hypothetical prote
12	99	11.7	219	2 B72291	hypothetical prote
13	97.5	11.5	614	2 A84152	hypothetical prote
14	97.5	11.5	1345	2 S46817	hypothetical prote
15	97	11.4	988	2 T14188	hypothetical prote
16	96.5	11.4	455	2 T05612	hypothetical prote
17	96	11.3	535	2 T37189	hypothetical prote
18	96	11.3	2500	2 G71609	hypothetical prote
19	95.5	11.3	644	2 T47835	hypothetical prote
20	95	11.2	348	2 T37271	cyclin II - human
21	95	11.2	1397	2 T10466	DNA topoisomerase
22	94.5	11.1	867	2 T27136	hypothetical prote
23	94.5	11.1	871	2 T27135	hypothetical prote
24	94.5	11.1	3724	2 T18427	hypothetical prote
25	94	11.1	210	2 T28771	hypothetical prote
26	93.5	11.0	645	2 E89883	conserved hypothet
27	93.5	11.0	649	2 S42488	dnak-type molecula
28	93	11.0	629	2 G96542	hypothetical prote
29	93	11.0	2464	1 QRMSP1	microtubule-associ

30	92.5	10.9	1888	2 T39009	hypothetical prote
31	92	10.8	688	2 A47705	triacylglycerol li
32	92	10.8	1202	2 S05362	probable DNA-direc
33	92	10.8	1875	2 S38173	myosin-like protei
34	91.5	10.8	650	2 D90093	heat shock protein
35	91	10.7	253	2 T32879	hypothetical prote
36	91	10.7	1087	2 T30330	gelatin-related p
37	90.5	10.7	443	2 S66040	serine-type D-Ala-
38	90.5	10.7	670	2 T28391	ORF MSV230 hypote
39	90.5	10.7	1332	2 S41552	probable transcrip
40	90.5	10.7	2401	2 T28676	rhoxy protein -
41	90	10.6	700	2 S67610	probable membrane
42	89.5	10.6	433	2 A89951	trigger factor (im
43	89.5	10.6	487	2 T45982	hypothetical prote
44	89.5	10.6	508	2 B81594	hypothetical prote
45	89.5	10.6	508	2 C72074	hypothetical prote

## ALIGNMENTS

```
RESULT 1
F95074
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: F95074
R:Rettein, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
son, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95074
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2140 <KIR>
A:Cross-references: GB:A005672; PIDN:NAK74791.1; PID:914972117; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0641

Query Match      100.0%; Score 848; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 6.5e-54;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVEKFLINKDTGEVSELKPRVVTYVTONGKMSSTIVSEBDPIIPYKGELEKGYOPDG 60
    |||
DB 1943 TTVEKFLINKDTGEVSELKPRVVTYVTONGKMSSTIVSEBDPIIPYKGELEKGYOPDG 2002

QY 61 WEISGFEKGDAGVYINISKDTPIKPVFKILEKKEENKPTFVSKKKDNPVNSQIN 120
    |||
DB 2003 WEISGFEKGDAGVYINISKDTPIKPVFKILEKKEENKPTFVSKKKDNPVNSQIN 2062

QY 121 ESHRKEDIQREHSHQSDSTVDVATVLDKNNISSKSTNNPNK 164
    |||
DB 2063 ESHRKEDIQREHSHQSDSTVDVATVLDKNNISSKSTNNPNK 2106

RESULT 2
A97942
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: A97942
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgess, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnen, S.; W
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:2143245; PMID:11544234
```

A:Accession: A97942  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <R>  
A:Cross-references: GB:AE007317; PID:AAK99365.1; PID:g15458138; GSPDB:GN00174  
C:Genetics:  
A:Gene: pTCA  
C:Keywords: hydrolyase; serine proteinase

Query Match	99.6%	Score 845;	DB 2;	Length 214;
Best Local Similarity	99.4%;	Pred. No. 1.1e-53;		
Matches 163; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0;

1947 TTYKFFLNKDTGEVSELKRRHTVITIQNGKEMSSIVSEEDPILPVKGELEKGYQFDG 60

2007 WEISGFEKKDAGIYNLSKDTFIKVEFKLIEKKEEENKPIFDVSKKKDNPQVNHSQLN 2066

2067 ESRKEDLDREDHSQKSDSTVDVTAFLDKNNISSKSTNNPK 2110

```

RESULT 3
CS497
laustrian - chicken
Alternate names: keratan sulfate proteoglycan

```

;Date: 07-Jul-1997 #sequence revision 12-Sep-1997 #text change 21-Jul-2000  
;Accession: JCS497; PC4334; S37561  
;Burg, M.A.; Cole, G.J.

Title: Claustrian, an adhesive neural keratan sulfate proteoglycan, is secreted by the developing mouse brain.  
 Reference number: JCS497; MUID:94157526; PMID:7906711  
 Accession: JCS497  
 Molecule type: mRNA

Accession: PC4334  
Cross-references: EMBL:X67778; NID:g406318; PIDN:CAA47988.1; PID:g406319  
Molecule type: protein

Experimental source: brain  
Comment: This protein inhibits neural cell adhesion and neurite outgrowth in chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate

Query March	13.1%	Score 111	DB 2	Length 1038
112,213,490/Binding site: carbonylrate (Asn)	(covalent)	#status predicted		
152,249,440,793,820/Binding site: chondroitin sulfate (Ser)	(covalent)	#status		

3 VKAEFLINKDTGVESELKPHRYTVIION----- 29

528 VKQAKLKQRTDSKESLKPAKTTTKQDCQKKNLKKHMSLSQIVQJLEKPKQLESKEKP 587

30 -----GKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGHEGKXDAGYINLSKDTF 83

84 IKPVFK-KIEEKKEEENKPTFDVSKKKDNPQVNHSQLNESHKREDLQRE-----EHSQKS 137

138 DSTKD V 143  
:: | : |

## RESULT 4

118467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2002  
C:Accession: 118467  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937

A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-558 <LAW>  
A:Cross-references: EMBL: AF006670. NID: 01407967. PDB: 1C610.1

C:genetics: 3  
A:Map position: 3  
A:Introns: 84/1; 160/1  
A:Note: C0465c

Query Match	13.0%;	Score 110;	DB 2;	Length 558;
Best Local Similarity	29.3%;	Pred. No. 1.1;		
Matches	54;	Conservative	23;	Mismatches 47;
				Indels 60;
				Gaps ,

Qy 37 IVSEEDFILPV-----KGLEKGYRPGDWELSGFEKK---DAGVYINLSKTETIKPV  
| : || | | | : || | : | : | :  
Db 60 ILGPEDDILYEYCISQLKQSKRK--KADGEEDKYLNAAKLKNLTGFIGNKKSDFIEEL

```

00  FAL--EALAB-----ENAFIVSA--ADUNVVAHSGUB-----SHAK
01  : : | | | : : : : :
118  LELLNKEEKEEHADTLNENK--TNDIKVKKNENINENNVNENKDISNKKKEHVSHQN
Db

```

177 EHNNVNLKKEKEYTDIGRDRKHKRSLSQSDSYKKRPFNKRKTSIER-SLSNKRKYDE

Db 236 KTNK 239

B71609  
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum

C/Accession: B71609  
R/gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.;  
Pierce, M.; Salberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.

A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: B71609  
A:status: preliminary; nucleic acid sequence not shown; translation not shown  
A:title: chromosome 2 sequence of the human malaria parasite Plasmodium falciparum

A,Residues: 1-665 <GAR>  
A,Cross-references: GB:AE001410; GB:AE001362; NID:93845245; PIDN:AACT1925.1  
A,Experimental source: clone 3D7

A;Gene: PFB0680W

Query Match	12.5%	Score 106;	DB 2;	Length 665;
Best Local Similarity	24.3%	Pred No 2	6;	

match	conservative	mismatches	indels	gaps
10 KDTGEVSELPKRVTVTQNGKEMSTVISEDPIPVYKGELEKQDFDWEI--SGF	35	40	36	

```

67  EGKDGAVINLSKDTPIKPVFKKIERKE-----EENKPTFD---VSKKKDNP
      |||      |||      |||      |||      |||      |||      |||      |||
      :|      :|      :|      :|      :|      :|      :|      :|

```

QY            113 ÖVNHSÖLNESHRKEDJÖR-EEHŠÖKSĐSTKVATVLDKNNISŠKSTNNPNK 164  
       : | :: : : | :::: ||| : :      ||      |:: : ||

Db 218 KVENKKSDDHKIEVKVKEHEDEBE-----DKKEKSKNKNDENK 261

## RESULT 6

D86432

hypochemical protein T518.14 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C:Accession: D86432

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: D86432

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-540 &lt;STO&gt;

A:Cross-references: GB:AE005172; NID:g4587525; PIDN:AAD25756.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

## Query Match

Best Local Similarity 12.3%; Score 104; DB 2; Length 540;

Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

1 TTVEFLNKDTGVESELKPHRVVTIONGKEMSSITVSEDFILPVYKGELEKYQFDG 60

Db 27 TTVKA-VVEETKVEDESKP-----EGVEKSAFKEESDPFADLKESEKK----- 70

Qy 61 WEISFEGKQAGVYINLSKQFIKPVFK--IEKKKEENKPTFDYSKKDNQVNS 117

Db 71 -ALSLKSLTEAIVDN---TLTKKKESSPMKEKKEEVKPEAEVKKKE--EAAEE 123

Qy 118 QLNESHKEDLQREHSQKSDSTKDTATVLDKNNISS 155

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Qy 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 124 KVEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

Db 113 NHIKSGYATNEIEIFLESCTLCKEITTAQTKNSYKRNINLKPREEEEEEEE 172

Qy 91 --IEKKEENKPPFDVSKKDNQVNSQLNESHKE-----DLQREHSQKSDSTK 141

Db 173 EEEQEEVEKPTISEEEBETPAVSEEEKKEEEEEETPAVSEEEKKEEEEEDEK 232

Qy 142 D 142

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

Db 233 D 233

Qy 233 D 233

RESULT 11  
T25911  
hypochemical protein T23B3.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25911  
R:Maggil, L.; Le, T.  
submitted to the EMBL Data Library, February 1997  
A:Description: The sequence of *C. elegans* cosmid T23B3.  
A:Reference number: Z20109  
A:Accession: T25911  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-211 <MAG>  
A:Cross-references: EMBL:U08309; PIDN:AAB4234.1; GSPDB:GN00019; CESP:T23B3.5  
A:Experimental source: strain Bristol N2; clone T23B3  
C:Genetics:  
A:Gene: CESP:T23B3.5  
A:Map position: 1  
A:introns: 30/2; 200/3

RESULT 13  
 AB4152  
 hypothetical protein BH4017 [imported] - *Bacillus halodurans* (strain C-125)  
 C:Species: *Bacillus halodurans*  
 C:Date: 01-Dec-2000 #sequence revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: AB4152  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
 A:Reference number: AB3650; MUID:20512582; PMID:11058132  
 A:Accession: AB4152  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-614 <STO>  
 A:Cross-references: GB:AP001520, GB:BA000004, NID:gi10176401, PIDN:BAB07736.1, GSPDB:GM00  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH4017

Query Match	11.5%;	Score 97.5;	DB 2;	Length 614;
Best Local Similarity	27.3%;	Pred. NO. 9.8;		

[illegible]

## 546817

hypothetical protein F9D16.270 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
C/Accession: T05612  
R/Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohnsels, J.; Mewes, H.W.; Mayer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A/Reference number: Z15419  
A/Accession: T05612  
A/Molecule type: DNA  
A/Residues: 1-456 <BEV>  
A/Cross-references: EMBL:AL053594  
A/Experimental source: cultivar Columbia; BAC clone F9D16  
C/Genetics:  
A/Map position: 4  
A/Intons: 110/3; 247/2; 282/3; 304/3; 361/3; 390/2; 418/3  
;Note: F9D16.270

**Best Local**

	Matches	40; Conservative	29; Mismatches	60; Indels	31; Gaps
Qy	10	KDTGEVSELPK	HRVTYTIQNGKEMSSIVSYSEEDFILPVPYKGGELBKGYQYQDGMIEISGFEG	68	
Db	245	KKKKKDPFLKPGHPVSALFYANRRRALREKNSVVEVAK	-----ITGEMKNLSD	296	
Qy	69	KKDAGY--VINLSKOTFLKPVFKKLEEKKEENKPTFDVSKKKDNPQVNHSSOLNESHKKE	126		
Db	297	KKKAPKEKAKKKKNTYLQ--AMEEKRTKEE-----ALSQKE-----EEELKLKHOB	345		
Qy	127	DLQREHSKSDSTKDTATATVLDKNNISSTKTVNN--PNK	164		
Db	346	ALQMLKKKEKTDN-----LIKKEKAKKKKKGNENVDNNK	378		

## T14188

hypothetical protein C02H7.1 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000  
 C:Accession: T37189  
 R:LeimDac, D.; Minx, M.  
 submitted to the EMBL Data Library, February 1996  
 A:Description: The sequence of C. elegans cosmid C02H7.  
 A:Reference number: Z20523  
 A:Accession: T37189  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-535 <LEI>  
 A:Cross-references: EMBL/U49945; PIDN:AA047924.1; GSPDB:GN00029; CESP:C02H7.1  
 A:Experimental source: strain Bristol N2; clone C02H7  
 C:Genetics:  
 A:Gene: CESP:C02H7.1  
 A:Map position: X

**BEST LOCAL**

Query Match	11.3%; Score 96; DB 2; Length 535;
-------------	------------------------------------

Best Local Similarity 21.3%; Pred. No. 11;  
Matches 35; Conservative 28; Mismatches 61; Indels 40; Gaps 5;

QY 11 DTGEVELKPHRTVTIIONGKEMSSITVSEDFILVYKGELEKGYQFDGMEISGPEGKK 70  
DB 77 DQSLKNVAAKTI---SGKDAET-----NKMQLMGTNATSNRN 116  
QY 71 DAGVYNLSKDTFIKPVFKIEKEEENKPTFDVSKKDNPOVNHSQLNESHRKEDLOR 130  
DB 117 GTC-----EKKKKKKVKEDKKGPBEKST---TKSSSKKETHKEKESKKSABE 166  
QY 131 EEHSQKSDSTK-----DVTATVLDKNNISSTSTNNPNK 164  
DB 167 KEKKKKSSSKERHKSSDRSEKSEKSEKKEKKEKSTDEKPK 210

## RESULT 18

G71609

Hypothetical protein PFB0650w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C&gt;Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C/Accession: G71609

R/Gardner, M.J.; Tetteelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
.; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A&gt;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; PMID:99021743; PMID:9804551

A:Accession: G71609

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2500 &lt;GAR&gt;

A:Cross-references: GB:AE001408; GB:AE001362; NID:g3845238; PIDN:AAC71919.1; PID:g384524

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0650w

Query Match 11.3%; Score 96; DB 2; Length 2500;  
Best Local Similarity 26.1%; Pred. No. 62;

Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;

QY 19 KPHRTVTIIONGKEMSSITVSEDFILVYKGELEKGYQFDGMEISGPEGKKDAGYVINTL 78  
DB 2173 KPYKIT---ENNKK-----NEGNEILKYSINENENKNNYDKQENNCILDKDTQCNVMT 2223  
QY 79 -----SKDTFIKPVFKIEKEEENKPTFDVSKKDNPOVNHSQLNE---SHRKED 127  
DB 2224 KEKNLNDKKSPSPNITKVKLEEEKSDDKRD---DKKNNTREKNLNDKKSPSPNITKVK 2280  
QY 128 LQREHSQKSDSTKDYATVLDKNNISSTSTNNPNK 164  
DB 2281 LEEEE---KSDDKRD-----DKKNNTREKNLNDKKSPSPNITKVK 2308

## RESULT 19

T47835

Hypothetical protein T209.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C&gt;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C/Accession: T47835

R/Wakatsuta, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
submitted to the Protein Sequence Database, February 2000

A:Reference number: Z24475

A:Accession: T47835

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-644 &lt;NYA&gt;

A:Cross-references: EMBL:AL138658

A:Experimental source: cultivar Columbia; BAC clone T209

C:Genetics:

A:Map position: 3

A:Insertions: 158/2; 329/3

A:Note: T209.90

Query Match 11.3%; Score 95.5; DB 2; Length 644;  
Best Local Similarity 23.4%; Pred. No. 14;  
Matches 34; Conservative 22; Mismatches 58; Indels 31; Gaps 4;

QY 28 QNGKEMSSITVSEB---DFILPVYKGELEKGYQFDGMEISGPEGKKDAGYVINTLSKDTF 83  
DB 519 ENSKTEKTKVADKKSVADFLRKIRKNSPQKK---ETTSKQKQNDGNV----- 565  
QY 84 IKPVFKIEKEEENKPTFDVSKKDNPOVNHSQLNESHRKEDLQREHSQKSDS--- 139  
DB 566 -----KENDHQKSDGNYKKNKSVPRELRSTGKKYKVENNNSSSKRKQ 615  
QY 140 TKDYATVLDKNNISSTSTNNPNK 164  
DB 616 TKETAEVATGKRGREGKDKQPRK 640

## RESULT 20

I37271

Cytich II - human

C:Species: Homo sapiens (man)

C&gt;Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000

C/Accession: I37271; S52774

R/Bess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.

Exp. Cell Res. 218, 174-182, 1995

A&gt;Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: chr

A:Reference number: I37271; PMID:9525491; PMID:7737358

A:Accession: I37271

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-348 &lt;HES&gt;

A:Cross-references: EMBL:Z46788; NID:g758586; PIDN:CAA86752.1; PID:g758587

Query Match 11.2%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 7.8;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 49 KSELEKGYQFDGMEISGPEGKKDAGYVINTLSKDTFIKPVFKIEKEEENKPTF---DV 105  
DB 205 ESBGEKG---GTEKDSKKKKKDS---KKGDSALIELQAVADEKDEGKKDANKGDE 256  
QY 106 SK--KKDNPOVNHSQLNE-----BSHRKEDLQREHSQKSDSTKD---VTATVLDKNNI 153  
DB 257 SKDAKDAKEIKKGGKKKPPSTSDSDKDVKE---SKDATYDAKVAKKDKTEKESA 313  
QY 154 SSK 156  
DB 314 DSK 316

## RESULT 21

T10466

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - malaria parasite (Plasmodium falc

C:Species: Plasmodium falciparum

C&gt;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000

C/Accession: T10466

R/Cheesman, S.J.

submitted to the EMBL Data Library, September 1995

A:Reference number: Z17031

A:Accession: T10466

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1397 &lt;CHE&gt;

A:Cross-references: EMBL:K79345; NID:g994807

C:Genetics:

A:Gene: TopoII

A:Map position: 14

C:Superfamily: eukaryotic type II DNA topoisomerase; phase T4 DNA topoisomerase (ATP-hy

C:Keywords: ATP; DNA binding; isomerase; nucleus

Query Match 11.2%; Score 95; DB 2; Length 1397;  
Best Local Similarity 23.8%; Pred. No. 38;  
Matches 36; Conservative 36; Mismatches 53; Indels 26; Gaps 6;

Query Match	11.1%;	Score 94.5;	DB 2;	Length 871;
Best Local Similarity	25.3%;	Pred. No. 24;		

Query Match	11.1%;	Score 94;	DB 2;	Length 210;
Best Local Similarity	28.6%;	Pred. No. 5.2;		





Qy 152 N-----ISSKSTNNPNK 164  
 Db 559 SKIAESSLQEIISASQIENSPTQ 582

## RESULT 29

GRMSPI

microtubule-associated protein MAP1B - mouse

N/Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei

C/Species: Mus musculus (house mouse)

C/Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text\_change 01-Sep-2000

C/Accession: S07549; S43387; A33645

R/Noble, M.; Lewis, S.A.; Cowan, N.J.

J. Cell Biol. 109, 3367-3376, 1989

A/Title: The microtubule binding domain of microtubule-associated protein MAP1B contains

A/Reference number: A33645; MUID:90094539; PMID:2480963

A/Accession: S07549

A/Molecule type: mRNA

A/Residues: 1-2464 &lt;NOB&gt;

A/Cross-references: EMBL:X51396; NID:G52999; PIDN:CAA35761.1; PID:G53000

R/Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.

Arch. Biochem. Biophys. 310, 428-432, 1994

A/Title: Binding of heat-shock protein 70 (hsp70) to tubulin.

A/Reference number: S44387; MUID:94234720; PMID:8179328

A/Accession: S44387

A/Status: preliminary

A/Molecule type: protein

A/Residues: 653-663, 'IC' &lt;SAN&gt;

C/Superfamily: microtubule-binding; phosphoprotein; tandem repeat

C/Keywords: microtubule binding; phosphoprotein; status experimental &lt;EMB&gt;

F:589-786/Domain: microtubule binding #status experimental &lt;EMB&gt;

F:589-592,639-642,649-652,655-656,660-663,668-671,674-677,679-682,683-686,687-690,691-69

R-K-E/D-X)

F:1861-2064/Region: 17-residue repeat

F:91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: PH

F:147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (cd

F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 11.0%; Score 93; DB 1; Length 2464;

Best Local Similarity 27.8%; Pred. No. 16+02;

Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;

Qy 4 KEFLINKDTGEVSELKPHRVTVITQNGKMSSTIVSEDFILPVYKGELEKGYQFGDMGI 63

Db 584 EKVYLVKDKPVKTESKP---SVT---EKVSS---KEEQ---SPV---KAQVA-----EK 623

Qy 64 SGFEGKADAGYVINKDTFIRPVFK-KIEEKKKEENKPTPDVSKKQDPVNHNSOLNS 122

Db 624 QATESKP-----KTKDKVKKKEIKTKLEEKKEE--KPKKEVKKEDKTPL---KKDK 672

Qy 123 HKEDLQRE-----EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164

Db 673 PKKEVKKKEIKKEIKKEERKELKEVKKETPLKDAKKEVKKKEKKEKPKK 728

## RESULT 30

T39009

hypothetical protein SPAC6B12.02c - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C/Accession: T39009

R/Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, August 1995

A/Reference number: Z21815

A/Accession: T39009

A/Status: preliminary; translated from GB/EMBL/DBDJ

A/Molecule type: DNA

A/Residues: 1-1888 &lt;GEN&gt;

A/Cross-references: EMBL:298531; PIDN:CAB11064.1; GSPDB:GN00066; SPDB:SPAC6B12.02c

C/Genetics:

A/Experimental source: strain 97zh-; cosmid c6B12

A/Map position: 1

C/Superfamily: Schizosaccharomyces hypothetical protein SPAC6B12.02c  
 Query Match 10.9%; Score 92.5; DB 2; Length 1888;  
 Best Local Similarity 23.1%; Pred. No. 81;  
 Matches 39; Conservative 25; Mismatches 76; Indels 29; Gaps 5;

Qy 1 TTVEFLINKDTGEVSELKPHRVTVITQNGKMSSTIVSE-EDFILPVYKGELEKGYQFD 59

Db 394 TSENFOLNVAANAVSTIPVTRTTTKKKRKRFKVEKLPDLISEY----- 442

Qy 60 GMEISGFEKADAGYVINKDTFIRPVFKKIEEKKKEENKPTPDVSKKQDPVNHNSOL 119

Db 443 -----GKAPKRLVAFARSSSHIP--KMIRRKQDSKKYFSDKESDROVIDYLS 492

Qy 120 NESRKEDELQREHSQKSDS-TKDYATVLDKN-----NISKSTNN 161

Db 493 DWYSGKHELVOOSHSHYKKRPSDKSVGNIFSVNSKSHVNINAKTPAANN 541

## RESULT 31

A47705

triacylglycerol lipase (EC 3.1.1.3) - Staphylococcus epidermidis

C/Species: Staphylococcus epidermidis

C/Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999

C/Accession: A47705

R/Farrell, A.M.; Foster, T.J.; Holland, K.T.

J. Gen. Microbiol. 139, 267-277, 1993

A/Title: Molecular analysis and expression of the lipase of Staphylococcus epidermidis.

A/Reference number: A47705; MUID:93171870; PMID:8436947

A/Contents: 9

A/Accession: A47705

A/Status: preliminary

A/Molecule type: DNA; protein

A/Residues: 1-688 &lt;FAR&gt;

A/Cross-references: GB:M95577; NID:G153021; PIDN:AAA19729.1; PID:G153022

A/Note: sequence extracted from NCBI backbone (NCBIN:125633)

C/Superfamily: Staphylococcus triacylglycerol lipase

C/Keywords: carboxylic ester hydrolase

Query Match 10.8%; Score 92; DB 2; Length 688;

Best Local Similarity 25.0%; Pred. No. 28;

Matches 40; Conservative 28; Mismatches 66; Indels 26; Gaps 7;

Qy 9 NKDTGEVSELKPHRVTVITQNGKMSSTIVSEDFILPVYKGELEKGYQFGDMGI 68

Db 68 NKNVNEKSNVN-----SITENESLHNETPKNEDLI-----QOQKDSQDNKSESVEQ 115

Qy 69 KKDAG-VYINKSKDTFIRPVFKIE-EKKEENKPTPDVSKKQDPV--NHSQLESNR 124

Db 116 NKENGAFVQNHSEB--KQOEQVLEKHAENNOTLHASKAQSNEYKTKPSQDLNTAA 172

Qy 125 KEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164

Db 173 KQEDSQKENLSKQDTQSSKTIDL-----RATQNOQSK 205

## RESULT 32

S05362

probable DNA-directed DNA polymerase (EC 2.7.7.7) - fungus (Ascobolus immerus) mitochondrion

C/Species: mitochondrion Ascobolus immerus

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C/Accession: S05362

R/Kempken, F.; Melhardt, F.; Besser, K.

Mol. Gen. Genet. 218, 523-530, 1989

A/Title: In organello replication and viral affinity of linear, extrachromosomal DNA of

A/Reference number: S05362; MUID:90066356; PMID:2571821

A/Accession: S05362

A/Molecule type: DNA

A/Residues: 1-1202 &lt;KEM&gt;

A/Cross-references: EMBL:X15982; NID:G2933; PIDN:CAA34106.1; PID:G1370212

C/Genetics:

A/Genetic code: SGC3

C:Superfamily: Ascobolus probable DNA-directed DNA polymerase  
C:Keywords: DNA binding; mitochondrion; nucleoside/transferase

Query Match 10.8%; Score 92; DB 1; Length 1202;

Best Local Similarity 22.5%; Pred. No. 53;  
Matches 42; Conservative 37; Mismatches 76; Indels 32; Gaps 9;

QY 6 FILNK-----DTGVSELSKPHRYVTIIONGKMSSTI---VSEEDF--ILPVYGELEK-- 54

DB 311 FVVNAKIKKPTGNVRSIGFNGVT-TLMDKTLIKITLAIPLERDHTVMSYDEGDI DESK 369

QY 55 ----GYQFDGWEISGEGKKAGYVINTSKPTFKVFKIE-----EKKEENK-PTF 103

DB 370 FPKGSLSPDKPKLTIEGTAVYTFPKKDIVVQINKKINFGDLPTXMDLSKMPNL 429

QY 104 DVSKKDNPQVNSQVNSHRRKDLQREHSQKSDSTKVTATVLDKNNI-----SSK 156

DB 430 KLNKDKTSGEIRMTIKNNQSYDI--IGHMINDENVTTFNRRAVNSIKIIFTVTDNM 487

QY 157 STNNPN 163

DB 488 GNTNDPN 494

#### RESULT 33

S38173  
myosin-like protein MLP1 - yeast (*Saccharomyces cerevisiae*)

N/Alternate names: protein YKR095W; protein YKR415

C/Species: *Saccharomyces cerevisiae*

C/Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 21-Jul-2000

R/Baladron, V.; Ballestra, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo, submitted to the Protein Sequence Database, March 1994

A/Accession: S38158

A/Reference number: S38158

A/Molecule type: DNA

A/Residues: 1-1875 <BAL>

A/Cross-references: EMBL:Z28320; NID:9486586; PID:9486587; MIPS:YKR095W

A/Experimental source: strain S288C

R/Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.

Mol. Gen. Genet. 237, 359-369, 1993

A/Title: A new yeast gene with a myosin-like heptad repeat structure.

A/Reference number: S31207; MUID:93247549; PMID:8483450

A/Accession: S31207

A/Molecule type: DNA

A/Residues: 1-300, 'A', 302-1875 <KOE>

A/Cross-references: EMBL:L01992; NID:g171958; PID:AAA34783.1; PID:g171959

C/Genetics:

A/Gene: SCD:MLP1

A/Cross-references: SGD:S0001803; MIPS:YKR095W

A/Map position: 11R

Query Match 10.8%; Score 92; DB 2; Length 1875;

Best Local Similarity 25.0%; Pred. No. 87;

Matches 48; Conservative 36; Mismatches 62; Indels 46; Gaps 10;

QY 7 ILNKDGEVSELKPHRYVTIIONGKMSSTI SEEDF-----ILPVYGE-----LEKGYOF 58

DB 660 LNK-----ELDIDVDSKSDISIKKGKSSRI LAEERKLLSNTLDLTAKENDDLRRKFDV 716

QY 59 DGEISGFGSKDA-----GYVINSKDTFKVPFKIIEK-----KEENK 100

DB 717 ----LQNTILKODSKTHETINRYVSCSKSIIVETELNLKEBOKLRVHLEKNLKOELNK 772

QY 101 PTFVSKKKDNPQVNSQVNSHRRKDLQREHSQKSDSTKVTATVLDKNNI-----TVLDKNN 151

DB 773 ----LSPEKDSLRINWTOQLQKEREDLLESTRSCCKKIDELDALSELKGETSQKDH 828

QY 152 NISSKSTNNPN 163

DB 829 HIKOLEEDNNSN 840

#### RESULT 34

D90093  
heat shock protein 70KD [imported] - *Galliardia theta* nucleomorph

C/Species: nucleomorph *Galliardia theta*

A/Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont

C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 15-Jun-2001

C/Accession: D90093

R/Douglas, S.; Zanner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei

Nature 410, 1091-1096, 2001

A/Title: The highly reduced genome of an enslaved algal nucleus.

A/Reference number: A99082; MUID:11323671; PMID:11323671

A/Accession: D90093

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-650 <DOU>

A/Cross-references: GB:AF165818; NID:g13794501; PID:AAK39876.1; GSPDB:GN00150

C/Genetics:

A/Map position: 1

A/Genome: nucleomorph

C/Suprafamily: heat shock protein 70

C:Keywords: nucleomorph

Query Match 10.8%; Score 91.5; DB 2; Length 650;

Best Local Similarity 23.6%; Pred. No. 29;

Matches 38; Conservative 19; Mismatches 51; Indels 53; Gaps 5;

QY 7 ILNKDGEVSELKPHRYVTIIONG---KMSSTI SEEDFILPVYKGELEKGYQFDGWEI 63

DB 492 ILNVSACDSTGSKSKITTTNDKGRLSKEIRMEVBAE---KYKNDEKTRQ----- 541

QY 64 SGPEKGDAGYVINSKDTFKVPFKIIEKEEEN-----KPTDVSKKKNPQVNSHQ 118

DB 542 -----KIEAKNNLNRYAVNINNTIRDEKDKDKIDENEKK 575

QY 119 LNES-----HRKEDLQREHSQKSDSTKVTATVLDK 150

DB 576 LBEKIRITLIEVENNEDLEKEDYERKEKELKNMSNPILSK 616

#### RESULT 35

TJ32879  
hypothetical protein C17F3.3 - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*

C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C/Accession: TJ32879

R/Gatung, S.; Scheet, P.

submitted to the EMBL Data Library, January 1998

A/Description: The sequence of *C. elegans* cosmid C17F3.

A/Reference number: Z21240

A/Accession: TJ32879

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-253 <GAT>

A/Cross-references: EMBL:AF043692; PIDN:AA97531.1; GSPDB:GN00019; CESP:C17F3.3

A/Experimental source: strain Bristol N2; clone C17F3

C/Genetics:

A/Gene: CESP:C17F3.3

A/Map position: 1

A/Introns: 41/1

Query Match 10.7%; Score 91; DB 2; Length 253;

Best Local Similarity 32.6%; Pred. No. 11;

Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;



A/Title: The genome of *Melanoplus sanguinipes* entomopoxvirus.  
A/Reference number: Z20484; MUID:99102612; PMID:9847359  
A/Accession: T28391  
A/Status: preliminary; translated from GB/EMBL/DBD  
A/Molecule type: DNA  
A/Residues: 1-670 <AFO>  
A/Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97713.1; PID:g4049753  
C/Genetics:  
A/Note: MSV230

Query Match 10.7%; Score 90.5; DB 2; Length 670;  
Best Local Similarity 22.5%; Pred. No. 35;  
Matches 36; Conservative 34; Mismatches 69; Indels 21; Gaps 5;

QY 25 VTIONGKMSSTI-----VSEEDFILPYKGLKGY-----QPDGWEISGP 66  
DB 33 VNFEEKQITSTLKFNNPDKTEMCGVSEKFFQVLNNKSASKEVDVSSIDESQNS 92  
QY 67 EGKKGAGYVNLKDTFIRKPKKIE-EKKEENKPTFDVSKKKNPOVNHSQL-NESHR 124  
DB 93 DSDSDGCVNIDESQNSDKVNINKLENSQNSDSKVNIDESQNSD-KVNINKLENSQ 151

QY 125 KEDLQREHESQKSDSTKDVATATVLDKNNISKSSTNNPNK 164  
DB 152 SDGKNVIDESQNSDKVNIDESQNSDKVNIDESQNSDK 191

RESULT 39  
S41552  
probable transcription factor SPT7 - yeast (*Saccharomyces cerevisiae*)  
N/Alternate names: protein YBR0739; protein YBR081C  
C/Species: *Saccharomyces cerevisiae*  
C/Date: 28-Jan-1994 #sequence\_revision 09-Sep-1994 #text\_change 20-Sep-1999  
A/Accession: S41552; S45946; S45948; S40800; S45478; S5485; S57716  
R/Gantheroff, L.; Dollard, C.; Tan, P.; Winston, F.  
A/Title: The EMBL Data Library, July 1993  
A/Reference number: S41552  
A/Molecule type: DNA  
A/Residues: 1-1332 <GAN>  
A/Cross-references: EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:g349190  
R/Steenma, H.Y.; van der Aart, Q.J.M.  
submitted to the Protein Sequence Database, August 1994  
A/Reference number: S45932  
A/Accession: S45946  
A/Molecule type: DNA  
A/Residues: 1-1332 <STE>  
A/Cross-references: EMBL:Z35950; NID:g536341; PIDN:CAA85026.1; PID:g536342; MIPS:YBR081C  
R/Andre, B.; Czjpluch, C.; Hein, C.; Jauniaux, J.C.; Urrestarazu, A.; Vissers, S.  
submitted to the Protein Sequence Database, August 1994  
A/Reference number: S45893  
A/Accession: S45948  
A/Molecule type: DNA  
A/Residues: 1-835 <AND>  
A/Cross-references: EMBL:Z35950; MIPS:YBR081C  
R/Haynes, S.R.; Dollard, C.; Winston, F.; Beck, S.; Trowdale, J.; David, I.B.  
Nucleic Acids Res. 20, 2603, 1992  
A/Title: The bromodomain: a conserved sequence found in human, *Drosophila* and yeast prot  
A/Reference number: S40800; MUID:92285152; PMID:1350857  
A/Accession: S40800  
A/Status: nucleic acid sequence not shown  
A/Molecule type: DNA  
A/Residues: 463-523 <HAY>  
A/Cross-references: EMBL:M67651; NID:g172683; PIDN:AAA35087.1; PID:g172684  
R/van der Aart, Q.J.M.; Barthe, C.; Dolignon, F.; Aigle, M.; Crouzet, M.; Steensma, H.Y.  
Yeast 10, 959-964, 1994  
A/Title: Sequence analysis of a 31 kb DNA fragment from the right arm of *Saccharomyces c*  
A/Reference number: S45462; MUID:95076715; PMID:7985423  
A/Accession: S45478  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-624; 'LRGKKRI', 633-1332 <VAN>  
A/Cross-references: EMBL:X76294

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993  
R/Gantheroff, L.J.; Dollard, C.; Tan, P.; Winston, F.  
Genetics 139, 523-536, 1995  
A/Title: The *Saccharomyces cerevisiae* SPT7 gene encodes a very acidic protein important  
A/Reference number: S54985; MUID:95229044; PMID:7713415  
A/Accession: S54985  
A/Status: nucleic acid sequence not shown  
A/Molecule type: DNA  
A/Residues: 1-1332 <GAM>  
A/Cross-references: EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:g349190  
R/van der Aart, Q.J.M.  
submitted to the EMBL Data Library, August 1995  
A/Reference number: S59702  
A/Accession: S59716  
A/Molecule type: DNA  
A/Residues: 1-1332 <VAM>  
A/Cross-references: EMBL:X76294; NID:g974203; PIDN:CAA53940.1; PID:e264674; PID:g558360  
A/Experimental source: strain S288C  
C/Genetics:  
A/Genes: SGD:SPT7  
A/Cross-references: SGD:S0000285; MIPS:YBR081C  
A/Map position: 2R  
C/Superfamily: unassigned bromodomain proteins; bromodomain homology  
C/Keywords: nucleus; transcription regulation  
F/466-521/Domain: bromodomain homology <BRO>

Query Match 10.7%; Score 90.5; DB 2; Length 1332;  
Best Local Similarity 24.5%; Pred. No. 76;  
Matches 39; Conservative 26; Mismatches 57; Indels 37; Gaps 8;

QY 25 VTIONGKMSSTI VSEEDFILPYKGLKGYQPDGWEISGEGKKGAGYVNLKDTFI 84  
DB 549 ITRIRRADLEKSI-----EDM-----EKDKYELDEEVEVAGSKRG-----LNMGAHMLA 594

QY 85 K--PYFKTIEKKKEENKPTPD-----VSKKD-----NPQVHSQLNEBH 123  
DB 595 KENGKVEKSDSKTYDEAPTDKLTSTVPRGEKCKTKTASTVTVAHNVKNEIKENG 654

QY 124 RKEDLQREHESQKSDSTKDVATATVLD-KNNISKSSTNN 160  
DB 655 KNEBQDMVEESKSTEDSSKDAADAKKDTEDGQDXTAEN 653

RESULT 40  
T28676  
rhophy protein - *Plasmodium yoelii* (fragment)  
C/Species: *Plasmodium yoelii*  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 01-Dec-2000  
A/Accession: T28676; A45521  
R/Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
Mol. Biochem. Parasitol. 76, 329-332, 1996  
A/Title: Comparison of two members of a multigene family coding for high-molecular mass  
A/Reference number: Z20507; MUID:97077455; PMID:8920022  
A/Accession: T28676  
A/Molecule type: DNA  
A/Status: preliminary; translated from GB/EMBL/DBD  
A/Residues: 1-2401 <SIN>  
A/Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AA41263.1  
R/Reen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.  
Mol. Biochem. Parasitol. 42, 241-246, 1990  
A/Title: Identification of the gene for a *Plasmodium yoelii* rhophy protein. Multiple co  
A/Reference number: A45521; MUID:91101660; PMID:2270106  
A/Accession: A45521  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 2260-2401 <KEE>  
A/Cross-references: GB:M34281

Query Match 10.7%; Score 90.5; DB 2; Length 2401;  
Best Local Similarity 27.4%; Pred. No. 15e+02;  
Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;

QY 26 TIIONGKMSSTI-----VSEEDFILPYKGLKGY-----QPDGWEISGEGKKGDA 72

```
Db      924 TSKNHEKISDIRKNSLKIIOFSESESYINDI-KKELEKNTVLESQNNNTDINGYLSKIEN 982
Qy      73 GVINLSKDOTPIKPVFKKIEEKKEENKPTFDVSKKKONPOVNHSQLNESHKEDLQREE 132
Db      983 IY--NILKLNKIKKIIDKVKEYTDEIEK-----NNKINAELSNSEKIITOLKENSLSKE 1035
Qy      133 HSQKSDSTKD-----VTATVLDKNISS 155
Db      1036 COSKIKSTIDDNVSECIKNITNLKTYIVNEKNININT 1072
```

Search completed: February 10, 2004, 10:58:35  
Job time : 12.8417 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 12.5638 Seconds  
(without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897  
Sequence: 1 KIVVDFPARNTTVEFLINK.....ATVLDKNNISSKSTTNPNK 174

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pirl:\*  
2: pirl:\*  
3: pirl:\*  
4: pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	2140	2 F95074	serine proteinase,
2	894	99.7	2144	2 A97942	metalloproteinase
3	114.5	12.8	1038	2 JCS497	claustrin - chick
4	110	12.3	558	2 T18467	hypothetical prote
5	109	12.2	665	2 B71609	hypothetical prote
6	104	11.6	540	2 D86432	hypothetical prote
7	103.5	11.5	325	2 T18283	hypothetical prote
8	102.5	11.4	385	2 T20410	hypothetical prote
9	102	11.4	622	2 A90570	lipoprotein (impor
10	101.5	11.3	312	2 G81339	probable membrane
11	101.5	11.3	335	2 T37189	hypothetical prote
12	101	11.3	700	2 S67610	probable membrane
13	100	11.1	219	2 B72291	hypothetical prote
14	100	11.1	1875	2 S38173	myosin-like protei
15	99	11.0	211	2 T25911	hypothetical prote
16	99	11.0	1397	2 T10466	DNA topoisomerase
17	97.5	10.9	614	2 A84152	hypothetical prote
18	97.5	10.9	1345	2 S46817	hypothetical prote
19	97	10.8	644	2 T47835	hypothetical prote
20	97	10.8	988	2 T14188	hypothetical prote
21	96.5	10.8	456	2 T05612	hypothetical prote
22	96.5	10.8	867	2 T27136	hypothetical prote
23	96.5	10.8	871	2 T27135	hypothetical prote
24	96	10.7	491	2 C97267	hypothetical prote
25	96	10.7	1939	2 T18372	repeat organellar
26	96	10.6	2500	2 G71609	hypothetical prote
27	95.5	10.6	3724	2 T18427	hypothetical prote
28	95	10.6	348	2 T37271	cyclin II - human
29	95	10.6	1202	1 S05362	probable DNA-direc

30	94.5	10.5	277	2 D70214	surface lipoprotei
31	94.5	10.5	670	2 T28391	ORF MSV230 hypoche
32	94	10.5	210	2 T28771	hypothetical prote
33	94	10.5	553	2 T15094	hypothetical prote
34	94	10.5	947	2 T03795	ornithine decarbox
35	93.5	10.4	456	2 T03045	hypothetical prote
36	93.5	10.4	456	2 B89883	conserved hypothet
37	93.5	10.4	649	2 S42488	dnak-type molecula
38	93	10.4	629	2 G96542	hypothetical prote
39	93	10.4	2464	1 Q8MSF1	microtubule-associ
40	92.5	10.3	443	2 S6040	serine-type D-Ala-
41	92.5	10.3	571	2 B64469	hypothetical prote
42	92.5	10.3	1702	2 A41859	IGA-specific metal
43	92.5	10.3	1888	2 T39009	hypothetical prote
44	92.5	10.3	1997	2 F71607	DNA helicase II BR
45	92	10.3	688	2 A47705	triacylglycerol 11

## ALIGNMENTS

```
RESULT 1
F95074
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: F95074
R:Retelish, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.O.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95074
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2140 <KIR>
A:Cross-references: GB:A005672; PIDN:AAK74791.1; PID:G14972117; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0641

Query Match      100.0%; Score 897; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 6.9e-57;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KIVVDFPARNTTVEFLINKDTGEVSELPKPRVVTIIONGKEMSTIYSEDFILPYK 60
DB      1933 KIVVDFPARNTTVEFLINKDTGEVSELPKPRVVTIIONGKEMSTIYSEDFILPYK 1992

QY      61 ELEKGYQDFGWEISGFGEKDGAGVYINLSKDTFLKPVFKIEEKEENKPTDVSXKD 120
DB      1993 ELEKGYQDFGWEISGFGEKDGAGVYINLSKDTFLKPVFKIEEKEENKPTDVSXKD 2052

QY      121 NPQVNSQLNSHREKEDIQREHSHQKSDSTQVATATVLDKNNISSKSTTNPNK 174
DB      2053 NPQVNSQLNSHREKEDIQREHSHQKSDSTQVATATVLDKNNISSKSTTNPNK 2106

RESULT 2
A97942
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: A97942
R:Hoeking, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.O.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, F.; McAnen, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Ballido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
```







A:Gene: MYPU 4650  
A:Genetic code: SGC3

Query Match 11.4%; Score 102; DB 2; Length 622;

Best Local Similarity 24.9%; Pred. No. 5.7;  
Matches 48; Conservative 35; Mismatches 66; Indels 44; Gaps 11;

QY 9 RNTTVEEFL-NDTGEVSEL-----KPHRV---TVTIQNG--KEM--STI--V 48  
DB 26 KNDASAKNFVYKSSDLISISKFPKRNNDKOKIIEGVITINSETKEIVKKEITLNL 65  
QY 49 SEEDFLLPYVKGELKGYQFDGWEISGFEGKKDAGVYINLSKDTFIKVPFKKIEEKEE 108  
DB 86 SEDNIIIFSLININNAKFPQDFE-----VSKDEKFKIKFOEINFSQTEQ 129  
QY 109 NKRTFVSKKKD---NPQVNSQLNES--HKEDIQREHSQKSDSTQVATVADKN- 161  
DB 130 -KITDNISKEDKKNPKKDNENSNNSSDQKNDELQKNNSDKLNVDKANKENSNS 188  
QY 162 NISSKSTNNPNK 174  
DB 189 NDSKEKNDENTNK 201

## RESULT 10

G81339  
probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C:Accession: G81339  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrett  
Nature 403, 665-668 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf  
A:Reference number: A81250; WUID:20150912; PMID:10688204

A:Accession: G81339  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-312 <PAR>  
A:Cross-references: GB:AL139076; GB:AL111168; NID:g6968128; PIDN:CAB72966.1; PID:g696814

A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:

A:Gene: Cj0692c

Query Match 11.3%; Score 101.5; DB 2; Length 312;

Best Local Similarity 25.0%; Pred. No. 2.8;  
Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;

QY 15 EFLINDTGEVSELKPHRVTVTIQNGKEMSSSTI---VSEDFLLPYK-----GELE 63  
DB 5 DFIKERQNTIRKMLKFSR---AINQKPLDDDLRDEISSDDILRRFPKKTTPKFLIED 61  
QY 64 KGYQFDGWEISGFEGKKDAGVYINL---SKDTFIKVPFKKIEEKEENKPT---FDVSK 117  
DB 62 EEEYSGHTKSKSNLYLKED---LINVKLEEKOSLAKIFSKKERRRRENNKTKKNLFSR 118  
QY 118 KKNP-----QVNSQLNESHRKEDIQREHSQKSDSTKV- -TATVLDKNNISK--STT 169  
DB 119 KKNKRNKNTQTKQIOTKSNQATTTQTKQEKELTNSIEKIQTETKIQKPLEIKLDVX 178  
QY 170 NNPN 173  
DB 179 NNPN 182

## RESULT 11

T37189  
hypothetical protein CO2H7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000

C:Accession: T37189  
R:leimDac, D.; Minx, M.  
submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid CO2H7.

A:Reference number: Z20523

A:Accession: T37189

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-535 <LEI>  
A:Cross-references: EMBL:U49945; PIDN:AAC47924.1; GSPDB:GN00029; CESP:CO2H7.1

A:Experimental source: strain Bristol N2; clone CO2H7

C:Genetics:

A:Gene: CESP:CO2H7.1

A:Map position: X

A:Intons: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 11.3%; Score 101.5; DB 2; Length 535;

Best Local Similarity 21.0%; Pred. No. 5.2;  
Matches 39; Conservative 31; Mismatches 71; Indels 45; Gaps 6;

QY 4 VKDFANTTVKEFIINK-----DTGEVSELKPHRVTVTIQNGKEMSSSTIVSEDFLLPY 58  
DB 55 IKSAGNDKNTKTAFLDKLIKIIDGSLKNVKAII---SGKDAEET----- 98  
QY 59 KGELEKGYQFDGWEISGFEGKKDAGVYINLSKDTFIKVPFKKIEEKEENKPTFDVSK 118  
DB 99 ---NKMQLMTNATSFNSRGTG-----EKKKKKKKKEDKKGBEEKST---TK 144  
QY 119 KNPQVNSQLNESHRKEDIQREHSQKSDSTK-----DYATVLDKNNISKST 168  
DB 145 RSKKETHEKRSKSKSAEKKKKSSSKENHKSDDRSEKSKSEKSKKKEKST 204  
QY 169 TNNPNK 174  
DB 205 TDEKPK 210

## RESULT 12

S67610  
probable membrane protein YDL074c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein D2483

C:Species: Saccharomyces cerevisiae  
C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002

C:Accession: S67610  
R:Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.  
submitted to the Protein Sequence Database, July 1996

A:Reference number: S67608

A:Accession: S67610

A:Molecule type: DNA

A:Residues: 1-700 <WAM>

A:Cross-references: EMBL:Z74122; NID:g1431087; PID:e253213; PID:g1431088; GSPDB:GN00004;

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:BRE1; MIPS:YDL074C

A:Cross-references: SGD:S0002232

A:Map position: 4L

C:Keywords: transmembrane protein

F:69-85/Domain: transmembrane #status predicted <TMM>

Query Match 11.3%; Score 101; DB 2; Length 700;

Best Local Similarity 22.4%; Pred. No. 7.7;  
Matches 46; Conservative 39; Mismatches 60; Indels 60; Gaps 10;

QY 5 KDFANTTVKEFIINK-----DTGEVSELKPHRVTVTIQ-----NGKEMSS 46  
DB 379 KDVNIRTRADDLKILAEKSKTEVLSDIQ-HAIDILKEQTKIDQSRNDTKSSST 437  
QY 47 IVSEDFLLPYVKGELKGYQFDGWEISGFEGSK-----DAGV 85  
DB 438 ---ODALIKETQ-DLEKGF---ELSDLTHKKSEIINNESVISKLVEKTKADQKX 488  
QY 86 INL-SKDTF---IKVFKKIEEKEENKPTFDVSKKKNPNQVNSQLNESHRKEDLORE 141  
DB 489 AAMRSKSDSLIKTKLSLSKSNEL-----ILQKSDRLQOKIGNKHQDLDSQ 541  
QY 142 EHSQKSDSTQVATVLDKNNISK 166

Db 542 NERRLIDSKTETLTILNTSTK 566

## RESULT 13

B72291

hypohectical protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: B72291

C:Author: K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.

C:Map position: 11R

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; PMID:99287316; PMID:10360571

A:Accession: B72291

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-219 &lt;ARN&gt;

A:Cross-references: GB:AE001771; GB:AE000512; NID:94981678; PIDN:AA036218.1; PID:9498168

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1142

Query Match 11.1%; Score 100; DB 2; Length 219;

Best Local Similarity 26.4%; Pred. No. 2.4;

Matches 47; Conservative 30; Mismatches 59; Indels 42; Gaps 10;

10 NTTVAFPLNDKGTGSEVSELPKRVTVTTQNGKMSSTIVSEEDFLLPYKGLKGYO-- 67

48 NLFFKEVPEPKEDFQFI--LIPFVVELYATDEKTTLPKKE-----VLGEESVYKDL 98

68 FDGMEISG-----FEKKDAGYVIN-LSKDTFIRPVFKLIEKKEENKPTFDVSKKD 120

99 FAGFEVGRGPTFFPKGKGGYLPGYVDKDNFK-ILKYVAQLKED---FQYTLKDD 153

121 NPQVNSQLNESHKED---LQREHSQKSDS-----TKDVTATYLDK 160

154 DPFVGEPLIETIF-KEDADFVLEKDNNAVKTVDVNEVRDRIVYTDSVDVAKTLQEK 210

## RESULT 14

S38173

myosin-like protein MLPI - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YKR095W; protein YKR415

C:Species: Saccharomyces cerevisiae

C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 21-Jul-2000

C:Accession: S38173; S40647; S31207

R:Baladron, V.; Baladron, J.P.G.; Bou, G.; del Rey, F.; Estebe, P.F.; Garcia-Cantalejo,

submitted to the Protein Sequence Database, March 1994

A:Reference number: S38158

A:Accession: S38173

A:Molecule type: DNA

A:Residues: 1-1875 &lt;BAL&gt;

A:Cross-references: EMBL:Z28320; NID:9486586; PID:9486587; MIPS:YKR095W

A:Experimental source: strain S288C

R:Bou, G.; Estebe, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Ji

Yeast 9, 1349-1354, 1993

A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chrom

A:Reference number: S40644; PMID:94205265; PMID:8154186

A:Accession: S40647

A:Molecule type: DNA

A:Residues: 1-1875 &lt;BOU&gt;

A:Cross-references: EMBL:X73541; NID:9450550; PIDN:CAA51948.1; PID:9450554

A:Experimental source: strain S288C

R:Koebling, R.; Nguyen, T.; Chen, E.Y.; Bortstein, D.

Mol. Gen. Genet. 237, 359-369, 1993

A:Title: A new yeast gene with a myosin-like heptad repeat structure.

A:Reference number: S31207; PMID:93247549; PMID:8483450

A:Accession: S31207

A:Molecule type: DNA

A:Residues: 1-300, 'A', 302-1875 &lt;KOE&gt;

A:Cross-references: EMBL:L01992; NID:G171958; PIDN:AA44783.1; PID:G171959

C:Genetics:

A:Gene: SGD:MLPI

A:Cross-references: SGD:S0001803; MIPS:YKR095W

A:Map position: 11R

Query Match 11.1%; Score 100; DB 2; Length 1875;

Best Local Similarity 24.4%; Pred. No. 28;

Matches 50; Conservative 38; Mismatches 71; Indels 46; Gaps 10;

4 VKDPAWTVKFLNDKGTGSEVSELPKRVTVTTQNGKMSSTIVSEEDF-----ILPVY 58

647 ISQTRSTENMSLNLK---EIQDYDSKSDISIKGKESRIIAERFYLNSNTLDLT 703

59 KGE---LEKGYQFQMGWISGFEKKDA-----GYVINSKDTFIRPVFKLIEK--- 104

704 KAENDQJLRKRDY---LQNTILKODSTHETLNEVYSCSKSISIVETELNLKEBOKLR 759

105 -----KEENKPTFDVSKKDNPOVNSQLNESHKEDL---QREHSQKSDSTKDV 153

760 VHLKNTLQELNK---LSPEKDSURLVWTLQTLQKEREDLBERTRKSCQKIDLEDA 815

154 TA-----TVLDKNNISSKSTNNPN 173

816 LSELKKTSSQKDHILKQLEEDNNSN 840

## RESULT 15

T25911

hypohectical protein T23B3.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T25911

R:Maggi, L.; Le, T.

submitted to the EMBL Data Library, February 1997

A:Description: The sequence of C. elegans cosmid T23B3.

A:Reference number: Z20109

A:Accession: T25911

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-211 &lt;MAG&gt;

A:Cross-references: EMBL:U88309; PIDN:AA84234.1; GSPDB:GN00019; CESP:T23B3.5

A:Experimental source: strain Bristol N2; clone T23B3

C:Genetics:

A:Gene: CESP:T23B3.5

A:Map position: 1

A:introns: 30/2; 200/3

Query Match 11.0%; Score 99; DB 2; Length 211;

Best Local Similarity 29.4%; Pred. No. 2.7;

Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

77 EGKDGAGYVINLSKDTFIRPVFKLIEKKEENKPTFDVSKKDNPOVNH-----SQLNS 132

69 EGKKGDKSKKXKDDGDKKEKKDKEDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDD 128

133 HKKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNNPNK 174

129 EKKDDKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEK 170

## RESULT 16

T10466

DNA topoisomerase (ATP-hydrolyzing) (EC 5.9.9.1.3) II - malaria parasite (Plasmodium falc

C:Species: Plasmodium falciparum

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000

C:Accession: T10466

R:Chesman, S.J.

submitted to the EMBL Data Library, September 1995

A:Reference number: Z17031

A:Accession: T10466

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A;Residues: 1-1397 <CHE>  
A;Cross-references: EMBL:X79345; NID:g9934807  
C;Genetics:  
A;Gene: TOP011  
A;Map position: 14  
C;Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hyd  
C;Keywords: ATP; DNA binding; isomerase; nucleus

Query Match 11.0%; Score 99; DB 2; Length 1397;  
Best Local Similarity 22.2%; Pred. No. 24;  
Matches 44; Conservative 45; Mismatches 67; Indels 42; Gaps 9;

QY 1 KIVVKDFAR-----NTTKEFILNKDGTGEVSELKPHRYVTITONGKMSSTIVSEE 51  
DB 1092 KVLVEELKRGYQPYKDKINIKKEEIFEGELDAAD-NPE-----DNEELIAGITVKKY 1144  
QY 52 DFLI--PVYKGELEKGYQPDGMEISGFEKKDAGYINISKOTFIKPVFKIIE----- 103  
DB 1145 DYLLSMPIFSLTLEK--VEDLLTQLEKERELEIINITIVETMWLKDIEKVEALEPQR 1201  
QY 104 -----KKEENKPTPDYSKKKNPQVNHSQLNESHKREDLQREHSQ---KSDTKDVA 155  
DB 1202 NVELSNREESNK--FKYARKQ---GPSMKKKKKKKLLSDSESGDTSSEPLVN 1254  
QY 156 TVLDKNNISSKSTNNPN 173  
DB 1255 TLNKKNTNKKTTSSNN 1272

## RESULT 17

A84152  
hypothetical protein BH4017 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: A84152  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: A83650; MIMD:20512582; PMID:11058132  
A;Accession: A84152  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-614 <STO>  
A;Cross-references: GB:AP001520; GB:BA000004; NID:g10176401; PIDN:BAB07736.1; GSPDB:GN00  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH4017

Query Match 10.9%; Score 97.5; DB 2; Length 614;  
Best Local Similarity 27.3%; Pred. No. 12;  
Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;

QY 26 SELKPHRYVTITONGKMSSTIVSE-----EDFILPVYKGELEKGYQPDGM---EISG 75  
DB 386 SEBKPE--TYTLQTALQM-TPIVNEYSPQTRBEFL-----ARKHQLDGMADVSKVY 435  
QY 76 FEKKDAGYINISKOTFIKPVFKIIEKK--EEENKPTPDVSKK---DNQVNHSQL 129  
DB 436 FAGNITASQL-----GKIEEKLDQKNNYITPDPFKKEVNVQSPKISTAL 482  
QY 130 NESHKREDLQREHSQKSDTKDVTATLD 159  
DB 483 GKVIQIVDLGRKKYHKGKEDLERLSKQIE 512

## RESULT 18

S46817  
hypothetical protein YHR080c - Yeast (Saccharomyces cerevisiae)  
C;Species: Saccharomyces cerevisiae  
C;Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 19-Apr-2002  
C;Accession: S46817  
R;Faville, T.  
submitted to the EMBL Data Library, June 1994

A;Description: The sequence of S. cerevisiae cosmid 9205.

A;Reference number: S46795  
A;Accession: S46817  
A;Molecule type: DNA  
A;Residues: 1-1345 <FAV>  
A;Cross-references: EMBL:U10556; NID:g500825; PIDN:AAB68895.1; PID:g500838; MIPS:YHR080C  
C;Genetics:  
A;Cross-references: SGD:S0001122  
A;Map position: 8R  
C;Keywords: transmembrane protein

Query Match 10.9%; Score 97.5; DB 2; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 29;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY 59 KGELEKGYQPDGMEISGFEKK-DAGYVINSKOTFIKPVFKIIEKKEENKPTFDYSK 117  
DB 1109 KCAIEKG-----SVGGQVSVDMYLSERDIT-----SRKSKKPVYKQVWK 1149  
QY 118 KKDNPQVNHSQLNESHKREDLQREHSQKSDTKDVTATLD--KNNISKSTN---NP 172  
DB 1150 SHDKRPFHSKVE-----QKSESRKSDNKOILTHILDFOVNNSSSEIFNMKLLSP 1201  
QY 173 NK 174  
DB 1202 QK 1203

## RESULT 19

T47835  
hypothetical protein T209.90 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C;Accession: T47835  
R;Yakutana, G.; Fattman, B.; Danner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
submitted to the Protein Sequence Database, February 2000  
A;Reference number: Z24475  
A;Accession: T47835  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-644 <NYA>  
A;Cross-references: EMBL:AL138658  
A;Experimental source: cultivar Columbia; BAC clone T209  
C;Genetics:  
A;Map position: 3  
A;Introns: 158/2; 329/3  
A;Note: T209.90

Query Match 10.8%; Score 97; DB 2; Length 644;  
Best Local Similarity 22.9%; Pred. No. 14;  
Matches 43; Conservative 29; Mismatches 72; Indels 44; Gaps 7;

QY 8 ARNTYKKEFILNKD--TGEV-SELKPHRYVTI-----IQNGKMSSTIVSEE--- 51  
DB 476 SRTSKELIIVAVAKDTKTKAKNNIKQDPTKTESSDDDDDEKEENSKTEKTVADKKSV 535  
QY 52 -PFIIPVYKGELEKGYQPDGMEISGFEKKDAGYINISKOTFIKPVFKIIEKKEENK 110  
DB 536 ADPLKRIKNSPOKKG-----ETSKNQKNDGNV-----KENDHQ 572  
QY 111 PTFDYSKKDNPQVNHSQLNESHKREDLQREHSQKSDS---TKDVTATLDKNNISSK 166  
DB 573 KSDGNVKKENSKVVPRELRSSTGKKVVEVNNSSSKSKQTKETAEVATGKGRREG 632  
QY 167 STTNPNK 174  
DB 633 KDDKQPRK 640

## RESULT 20

T1188  
hypothetical protein T28D5.30 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)

RESULT 22  
T27136  
hypotheical protein Y53C12B.3a - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*

RESULT 24  
C97267  
hypothetical protein CAC2985 [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #ext\_change 14-Sep-2001  
C:Accession: C97267  
R:Moiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C1C  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: C97267  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-491 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK80926.1; PID:G15026040; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2985

Query Match 10.7%; Score 96; DB 2; Length 491;  
Best Local Similarity 24.6%; Pred. No. 12;  
Matches 51; Conservative 33; Mismatches 69; Indels 54; Gaps 11;

QY 1 KIVVDFPANTVKEFIINKDTGEVSELKPHRYVTITONGKMSSTIVSEEDFILPVYKG 60  
DB 77 KINIDVYGRSEISREF-----EESDVAGSGIADVGVGASISOPF---EDNIYVWSG 126  
QY 61 ELEKGVQPDGWEISG-----FEGKKDAGVYINLSKOTFIKPV---FKIE-----E 103  
DB 127 -----FNASDVSDSKVSPVVEKKEAPV---KREBVKPVKKODKVEKESTTD 175  
QY 104 KKEENKPTFD-----VSKKON-POVNSOLNE-----SHRKEDLOREHSOKSD----- 148  
DB 176 KKENLNIRNIFDKYEEKIEKEKKNIEKKNHEIEKKEVNTKKEDPLKEKEETNVDDKKEPR 235  
QY 149 -STKDVATVLDKNNISSKSTTNPNK 174  
DB 236 VEEKETEVEVKEIEVKKEETKEESR 262

## RESULT 25

T18372  
repeat organelar protein - Plasmodium chabaudi  
C:Species: Plasmodium chabaudi  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #ext\_change 15-Oct-1999  
C:Accession: T18372  
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.  
Mol. Biochem. Parasitol. 94, 185-196, 1998  
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted spe  
A:Reference number: Z18922; MUID:98418765; PMID:9747969  
A:Accession: T18372  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1939 <WER>  
A:Cross-references: EMBL:U43145; NID:G1151157; PID:G1151158; PIDN:AA63403.1

Query Match 10.7%; Score 96; DB 2; Length 1939;  
Best Local Similarity 22.6%; Pred. No. 57;  
Matches 47; Conservative 46; Mismatches 73; Indels 42; Gaps 10;

QY 1 KIVVDFPANTVKEFIINKDTGEVSELKPHRYVTITONGKMSSTIVSEEDFIL 55  
DB 339 KLEIKDLSQSLVEKEKEIFEIKNEYDKINMKKEKLSI---NDKIDNTVLSSEKIN 395  
QY 56 PUYK-----GELEKYQPDGWEISGFEKKDAGY-----VINLSKOTFIKPVK 99  
DB 396 KILKEKETELNEIHKYKYLEIETIKNELNEKEKELEKKNKAHTVEVTNLTKE--IKLLEK 453  
QY 100 KIEEKE-----EENKPTFDVSKKONPOVNSOLNES---HRKEDLOREHSOKSD 148  
DB 454 KTEBDAGEKKNELNEINLQSLKANKKDIKNENTELNDKISSLSNEVUINLKDQTLGN 513  
QY 149 STKDVATVLDKNNISSKSTTNPNK 174  
DB 514 DIKTLDLI---NNLKNINTSDNKNK 538

## RESULT 26

G71609  
hypothetical protein PFB0650w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #ext\_change 21-Jul-2000  
C:Accession: G71609  
R:Gardner, M.J.; Tettein, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Pertes, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: G71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2500 <GAR>  
A:Cross-references: GB:AE001408; GB:AE001362; NID:G3845238; PIDN:AACT1919.1; PID:G384524  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0650w

Query Match 10.7%; Score 96; DB 2; Length 2500;  
Best Local Similarity 26.1%; Pred. No. 77;  
Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;

QY 29 KHRVTVITONGKMSSTIVSEEDFILPVYKGELEKGVQPDGWEISGFEKKDAGVINTL 88  
DB 2173 KPYKIT---ENNK-----NEGNEILKYSIENEKKNYDKQENBCLDKDTQCNVNT 2223  
QY 89 -----SKOTFIKPVFKIEKEEENKPTFDVSKKONPOVNSOLNE-----SHRKED 137  
DB 2224 KKKNNLNDKKSFPSPNIVKYLEEESKSDKRD---DKKNDTRKNNLNDKKSFPSPNIVK 2280  
QY 138 LOREHSOKSDSTKDVATVLDKNNISSKSTTNPNK 174  
DB 2281 LEEER---KSDDKRD-----DKKNDTRKNNLNDK 2308

## RESULT 27

T18427  
hypothetical protein C0335C - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #ext\_change 09-Jun-2000  
C:Accession: T18427  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, August 1997  
A:Reference number: Z18935  
A:Accession: T18427  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3724 <LAW>  
A:Cross-references: EMBL:Z98547; NID:E1325376; PID:E1325379; PIDN:CAB11104.1  
A:Genetics:  
A:Introns: 307/1; 1545/2  
A>Note: C0335C

Query Match 10.6%; Score 95.5; DB 2; Length 3724;  
Best Local Similarity 19.1%; Pred. No. 1.3e+02;  
Matches 40; Conservative 36; Mismatches 66; Indels 67; Gaps 7;

QY 1 KIVVDFPANTVKEFIINKDTGEVSELK----- 29  
DB 982 KILNDKMKKKEQEFKIDETGSIQSHKIKYKNGEKKHDKNNNEKNILYDENVYV 1041  
QY 30 ---PHRYVTITONGKMSSTIVSEEDFILPVYKGELEKGVQPDGWEISGFEKKDAGYVI 86  
DB 1042 LYSCHKIEDQIDHSIQNICDENN-IEQINEENSKKGVRIISGIDM---ENKND----- 1092  
QY 87 NLSKOTFIKPVFKIEKEEENKPTFDVSKKONPOVNH-SQLNEHSRKEDLOREHSQ 145  
DB 1093 -----MENKNDMEKN--DMEKKNDIEKNDMEKKNDEKKNDEKKNDE 1136

Oy 146 KSDSTKDVATVLDKNISKSTNNPK 174  
 Db 1137 KKNIME-----NENNENKSDIENENK 1158

## RESULT 28

1137271  
 Cyllicin II - human  
 C:Species: Homo sapiens (man)  
 C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000  
 C:Accession: 137271; S52774  
 R:Hees, H.; Heid, H.; Zimelmann, R.; Franke, W.W.  
 Exp. Cell Res. 218, 174-182, 1995  
 A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the  
 A:Reference number: 137271; MUID:9525491; PMID:7737358  
 A:Accession: 137271  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-348 <HE>  
 A:Cross-references: EMBL:Z46788; NID:9758586; PIDN:CAA6752.1; PID:9758587

Query Match 10.6%; Score 95; DB 2; Length 348;  
 Best Local Similarity 30.9%; Pred. No. 9.4;  
 Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

Oy 59 KGELEKGYQFDGWEISGFEKKDAGYVNLKDTIKPVFKKIEKKBEENKPTF---DV 115  
 Db 205 ESEGEKG-----GTEKDSKKKKDS---KKGDSAKIELQAVKADKKDEGDKKANKGDE 256  
 Oy 116 SK--KKDPOVNSOLN-----ESHKEDLQREHSQKSTKD---VTATVLDKNNI 163  
 Db 257 SKDAKKAKKIKKKKKKKSTSDSKDVKK---SKDADKAKKAVAKKTEKESA 313

Oy 164 SSK 166  
 Db 314 DSK 316

## RESULT 29

S05362  
 Probable DNA-directed DNA polymerase (EC 2.7.7.7) - fungus (Ascobotus immerus) mitochon  
 C:Species: mitochondrion Ascobotus immerus  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
 C:Accession: S05362  
 R:Kempken, F.; Meinhardt, F.; Esser, K.  
 Mol. Gen. Genet. 218, 523-530, 1989  
 A:Title: In organello replication and viral affinity of linear, extrachromosomal DNA of  
 A:Reference number: S05362; MUID:90066356; PMID:2573821  
 A:Accession: S05362  
 A:Molecule type: DNA  
 A:Residues: 1-1202 <KEM>  
 A:Cross-references: EMBL:X15982; NID:g2933; PIDN:CAA34106.1; PID:g1370212  
 C:Genetics:

A:Genome: mitochondrion  
 A:Genetic code: SGC3  
 C:Superfamily: Ascobotus probable DNA-directed DNA polymerase  
 C:Keywords: DNA binding; mitochondrion; nucleotidyltransferase

Query Match 10.6%; Score 95; DB 1; Length 1202;  
 Best Local Similarity 22.8%; Pred. No. 39; Indels 32; Gaps 9;  
 Matches 44; Conservative 37; Mismatches 80; Indels 32; Gaps 9;

Oy 10 NTTVEFLINK-----DTGEVSELKPHRYVTITQNGKMSSTI---VSEEDF--ILPVYKG 60  
 Db 305 NPTQFVFNNAKIKPTGNVRSIGQNT-TLTDKETILIKLAIPLERDITWVSDEG 363  
 Oy 61 ELK-----GYQFGWEISGFEKKDAGYVNLKDTIKPVFKKIE-----EKSEF 108  
 Db 364 DIDESKPRKGLSPDFKPLKTEGKRYANYTPPIKDIIVVCKINFNGLDKPTMDL 423  
 Oy 109 NK-PFFDVSKKKDNQVNSQLNESHKEDLQREHSQKSTKDVTATVLDKNNI----- 163  
 Db 424 SKMPLKLNKDKTSGEIRKTIKKNKNNOSYDI--IGHMIINDGENVITTRRAVDNIIKIF 481

Oy 164 ---SSKSTNNPN 173  
 Db 482 TVTDSMGNTNDPN 494

## RESULT 30

D70214  
 Surfactant lipoprotein P27 homolog - Lyme disease spirochete plasmid A/tp54  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
 C:Accession: D70214  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lachigra, R.; White  
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
 A:Reference number: A70100; MUID:98065943; PMID:9403685  
 A:Accession: D70214  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-277 <KLE>  
 A:Cross-references: GB:AC000790; NID:g2690224; PIDN:AAC66239.1; PID:g2690238; TIGR:BBA60  
 A:Experimental source: strain B31  
 C:Genetics:  
 A:Genome: plasmid

Query Match 10.5%; Score 94.5; DB 2; Length 277;  
 Best Local Similarity 20.9%; Pred. No. 7.8;  
 Matches 34; Conservative 28; Mismatches 62; Indels 39; Gaps 4;

Oy 4 VKDFARNTTVKEFLINKDTGEVSELKPHRYVTITQNGKMSSTIYSEDFILPVYKGELE 63  
 Db 142 LKNFEKIKNIENFPQNDLFLVTLKDKNNNTINIMLPNDIQKPDYILKDKTKIK 201  
 Oy 64 KGYQFDGWEISGFEKKDAGYVNLKDTIKPVFKKIEKKBEENKPTFVSKKKDNPQ 123  
 Db 202 KG-----TGEK-----YINPIYR-----FOIKKKDYHS 225

Oy 124 VNSQLNESHKEDLQREHSQKSTKDVA---TVLDKNNI 163  
 Db 226 IDYKVTISSEKTIEDLLPHQVQFMKNFKIITITIDNLN 268

## RESULT 31

T28391  
 ORF MSV230 hypothetical protein - Melanoplus sanguinipes entomopoxvirus  
 C:Species: Melanoplus sanguinipes entomopoxvirus  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T28391  
 R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
 J. Virol. 73, 533-552, 1999  
 A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
 A:Reference number: Z20484; MUID:99102612; PMID:9847359  
 A:Accession: T28391  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-670 <AFQ>  
 A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97713.1; PID:g4049753  
 C:Genetics:  
 A:Note: MSV230

Query Match 10.5%; Score 94.5; DB 2; Length 670;  
 Best Local Similarity 21.9%; Pred. No. 22;  
 Matches 42; Conservative 39; Mismatches 82; Indels 29; Gaps 6;

Oy 3 VKDFARNTTVKEFLINKDTGEVSELKPHRYVTITQNGKMSSTI-----VS 49  
 Db 9 ILLDKVINMKIEKFI-----KLKKYQNVNVPFEKKQIISTLTKENNDFKTEMCGVS 60  
 Oy 50 EEDFILPVYKGELEKGY-----QFDGWEISGFEKKDAGYVNLKDTIKPVFKKIE-E 103

30 / VNUXBEI1VQ--VNUXGVBSE1GQINFEEENUEGEINNESS--EEEDGVLN 301



Query Match 10.4%; Score 93.5; DB 2; Length 456;  
 Best Local Similarity 20.3%; Pred. No. 16;  
 Matches 41; Conservative 39; Mismatches 83; Indels 39; Gaps 5;

3 VKQDPAATVTYKFLINKDTGSELSKPHRVVTIIONGKEMSSITVSEEDFILPVYKSEL 62  
 70 VVQSSINSTPIRTEL--KDTPIYEE-TPIKRIITVTNKTIVSSISGNGNRNRIYDDL 126  
 63 EKGYPDGMWISGFEKKDAGVYINLSKDTFI---KPVFKIEKKE---EENKPT 112  
 127 FDDDDYKSPTRKPFQGEKDEDIRLIPKSSNIGSSKYKVLTRVENEKIKIHIDQKES 186  
 113 FDSKSKDNPQ-----VNHQSLSNHRKEDLQREHSGK 146  
 187 IVNDRKKNPFEREKPKDKEDKVKPPSLKEIKENKGIIDHENEEDKRELKFLQLOK 246  
 147 SDSLDTATVATVDKNNISSKPT 168  
 247 QYPLRDIPTIRSEYKSMKT 268

## RESULT 36

conserved hypothetical protein SA0976 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: E89883  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc, ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1235-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: E89883  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-645 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13700929; PIDN:BA042225.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: SA0976

Query Match 10.4%; Score 93.5; DB 2; Length 645;  
 Best Local Similarity 21.9%; Pred. No. 24;  
 Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

11 TTVEKFLINKDTGE-VSELKPHRVTV--TIONGKEMSSITVSEEDFILPVYKGELE-KGY 66  
 327 SAITEFQNVQPTNEKMTLDQTKYVYVESVENNESMTDFVKH-----PIKTMANGKKT 381  
 67 QF-----DGMWISGFEKK-----D 81  
 382 MWMTETNDYKDFWVEGGRVATISKDAKNRTIIFPYVEKGTLYDAIVKHYVTIDYD 441  
 82 AGYVYINL-SKDTFIPVFKIEKEEENKPTFDV-----SKKKNPOVNHQSLSNHRK 135  
 442 GQYHRYIVDKFAFTANTDKSNKKEODNSAKKEKTPATPSKPTPSVPEKESQKODSQKD 501  
 136 EDLQ-----REHSQSDSTKQVT-ATVLDKNNISSKSTNNNNK 174  
 502 DKQPLSVKEKNDASSGCKDTPATKPTKGEVSSSTI--PTK 543

## RESULT 37

SA2488

dnak-type molecular chaperone hsp70 - Pyrenomonas salina nucleomorph

N:Alternate names: heat shock protein 70

C:Species: nucleomorph Pyrenomonas salina

A:Note: A nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont

C:Date: 13-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 20-Sep-1999

C:Accession: S42488; S45576

R:Reising, S.; Hofmann, C.J.B.  
 Submitted to the EMBL Data Library, March 1993

A:Description: Smallest known eukaryotic genomes encode a protein gene: towards understa  
 A:Reference number: S42488  
 A:Accession: S42488  
 A:Molecule type: DNA  
 A:Residues: 1-649 <RES>  
 A:Cross-references: EMBL:X72621; NID:9461335; PIDN:CAA51197.1; PID:9461336  
 R:Hofmann, C.J.B.; Reising, S.A.; Haebler, M.M.; Martin, W.F.; Mueller, S.B.; Couch, J.; Mol. Gen. Genet. 243, 600-604, 1994  
 A:Title: The smallest known eukaryotic genomes encode a protein gene: towards an underst  
 A:Reference number: S45576; MUID:94268506; PMID:8208251  
 A:Accession: S45576  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-24,'Q',26-91,'H',93-578 <HOB>  
 A:Cross-references: EMBL:X72621; NID:9461335  
 A:Note: the sequence is revised in Genbank entry PSHSP70, release 111.0, (PIDN:CAA51197.  
 C:Genetics:  
 A:Gene: hsp70  
 A:Genome: nucleomorph  
 C:Function:  
 A:Description: involved in protein folding and assembling/disassembling of protein comp  
 C:Superfamily: heat shock protein 70  
 C:Keywords: ATP; molecular chaperone; nucleomorph

Query Match 10.4%; Score 93.5; DB 2; Length 649;  
 Best Local Similarity 24.5%; Pred. No. 25;  
 Matches 38; Conservative 26; Mismatches 50; Indels 41; Gaps 7;

17 ILNKDTGVSSELKPHRVTVTIIONG--KEMSTVSEEDFILPVYKGELEKGYQFDGWEI 73  
 491 ILNVSASAKSKGSKKXKTIITNDKGRLSKEELERVAEAE-----KKTEDEK----- 537  
 74 SGFEKKDA-----GYVINLS--KDTFIKPVFKIEKEEENKPTFDVSKKKNPOVN 125  
 538 --LDKLEAKNSLEAYVAINRTVAD-----EKLEKIOEDPKSIEEKVKE----- 582  
 126 HSQLSNHRKEDLQREHSGKSDSTKQVTATVLOK 160  
 583 --VLEFIETNEDLEKEVEKEKELKNPANTISK 615

## RESULT 38

G96542

hypothetical protein F1706.14 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: G96542

R:Thelloglis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matli, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G96542

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-629 <STO>

A:Cross-references: GB:AB005173; NID:g11054631; PIDN:AA027876.1; GSPDB:GN00141

C:Genetics:

A:Gene: F1706.14

A:Map position: 1

Query Match 10.4%; Score 93; DB 2; Length 629;  
 Best Local Similarity 19.6%; Pred. No. 26;  
 Matches 40; Conservative 39; Mismatches 73; Indels 52; Gaps 7;

11 TTVEKFLINKDTGEV-----SELKPHRVTVTIIONGKEMSSITVSEEDFI 54

Db 391 TTAK--MSSSTAELVOLPAKGVGKMDOKVSGEPMHLETKAKTDSAMEQIVAEADVAM 447  
 QY 55 LPPVKEELKGVQFDMWEISGFECKKDGAVINLSKDTIKPVEKKEEKE-----106  
 Db 448 NPIVEKAMEWEAEAGAAINPIVEADG-----ANNPIVEKAMSQIVAEADAINQAVD 501  
 QY 107 ---EENKPFVDVSKKKDNPO--VNSQLNESHKREDLQREHESQKSDTKDVATVADLN 161  
 Db 502 ANRQOTAPFGNDAAEDDSEPSHS---FTLNPLEKKEKVRKQATERSVADQODKN 558  
 QY 162 N-----ISKSTNNPNK 174  
 Db 559 SKTIAESSLOEFISASQTENSPFO 582

RESULT 39  
 ORMSPI  
 microtubule-associated protein MAP1B - mouse  
 N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei  
 C:Species: Mus musculus (house mouse)  
 C>Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #ext\_change 01-Sep-2000  
 C:Accession: S07549, S44387, A33645  
 R:Noble, M.; Lewis, S.A.; Cowan, N.J.  
 J. Cell Biol. 109, 3367-3376, 1989  
 A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains  
 A:Reference number: A33645, MUID:90094539, PMID:2480963  
 A:Accession: S07549  
 A:Molecule type: mRNA  
 A:Residues: 1-2464 <NOB>  
 A:Cross-references: EMBL:X51396; NID:952999; PIDN:CA53761.1; PID:953000  
 R:Sanchez, C.; Paducci, R.; Paciucci, R.; Zabala, J.C.; Avila, J.  
 Arch. Biochem. Biophys. 310, 482-492, 1994  
 A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.  
 A:Reference number: S44387, MUID:94234720, PMID:8179328  
 A:Accession: S44387  
 A:Molecule type: protein  
 A:Status: preliminary  
 A:Residues: 653-663, 'IC' <SAN>  
 C:Superfamily: microtubule-associated protein MAP1B  
 C:Keywords: microtubule binding; phosphoprotein; tandem repeat  
 F:589-186/Domain: microtubule binding #status experimental <MTB>  
 F:589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-69  
 R-X/E/D-X)  
 F:1861-2064/Region: 17-residue repeats  
 F:116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: ph  
 F:447,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (Co  
 F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 10.4%; Score 93; DB 1; Length 2464;  
 Best local Similarity 27.8%; Pred. No. 1,2e+02;  
 Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;

QY 14 KEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEI 73  
 Db 584 EKVLVKKDKPVKTESKP---SVT---EKEVSS---KEEQ--SPV-KAEVA-----EK 623  
 QY 74 SGEFGKDGAGYVNIISKQPIKRVFK-KIEEKEENKPPFDVSKKKDNQVNHSLNES 132  
 Db 624 QATESKP-----KVTKDKVKKKEITKLEKEE--KPKKEVKKEDKTP--KDEK 672  
 QY 133 HRKEDLQRE-----EHSQKSDTKDVATVADKNISKSSTNNPNK 174  
 Db 673 PRKEVKEIKKEIKKEERKELEKGVKKEKTPPLDKAKVKEKEKEKKEKPKK 728

RESULT 40  
 S66040  
 Serine-type D-Ala-D-Ala carboxypeptidase (EC 3.4.16.4) daaa - Bacillus subtilis  
 N:Alternate names: penicillin-binding protein 5  
 C:Species: Bacillus subtilis  
 C>Date: 28-Oct-1996 #sequence revision 13-Mar-1997 #ext\_change 20-Jun-2000  
 C:Accession: S66040, I39830, A92275, A92307, B61335, D69612, A23307  
 R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.

DNA Res. 1, 1-14, 1994  
 A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom  
 A:Reference number: S65967, MUID:96051385, PMID:7584024  
 A:Accession: S66040  
 A:Molecule type: DNA  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Residues: 1-443 <OGA>  
 A:Cross-references: EMBL:D26185; NID:9467326; PIDN:BA05246.1; PID:9467400  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993  
 R:Rodd, J.A.; Roberts, A.N.; Johnstone, K.; Piggett, P.J.; Winter, G.; Ellar, D.J.  
 J. Bacteriol. 167, 257-264, 1986  
 A:Title: Reduced heat resistance of mutant spores after cloning and mutagenesis of the B  
 A:Reference number: I39830, MUID:86250602, PMID:3087956  
 A:Accession: I39830  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 48-226, 'O', 228-443 <RES>  
 A:Cross-references: GB:M13766; NID:9142816; PIDN:AAA22375.1; PID:9142817  
 R:Maxam, D.J.; Strominger, J.L.  
 J. Biol. Chem. 255, 3964-3976, 1980  
 A:Title: Sequence of active site peptides from the penicillin-sensitive D-alanine carbox  
 A:Reference number: A92275, MUID:80182289, PMID:6768745  
 A:Accession: A92275  
 A:Molecule type: protein  
 A:Residues: 32-95, 'X', 97-98, 'XOX', 102 <MAX>  
 R:Maxam, D.J.; Strominger, J.L.  
 J. Biol. Chem. 255, 2067-2077, 1981  
 A:Title: Primary structure of the COOH-terminal membranous segment of a penicillin-sensi  
 A:Reference number: A92307, MUID:8117303, PMID:6780559  
 A:Accession: A92307  
 A:Molecule type: protein  
 A:Residues: 414-443 <MA2>  
 R:Yocum, R.R.; Maxam, D.J.; Rasmussen, J.R.; Strominger, J.L.  
 Proc. Natl. Acad. Sci. U.S.A. 76, 2730-2734, 1979  
 A:Title: Mechanism of penicillin action: penicillin and substrate bind covalently to the  
 A:Reference number: A61335, MUID:79223865, PMID:111240  
 A:Accession: B61335  
 A:Molecule type: protein  
 A:Residues: 55-68 <YOC>  
 R:Kunze, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allioni, G.; Azevedo, V.; Berter  
 C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
 A.; Ehrlich, S.D.; Emerson, P.T.; Eutlan, K.D.; Eyring, J.; Fabre, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
 lech, J.; Harwood, C.R.; Hensel, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F  
 Koester, P.; Konigstein, G.; Krogh, A.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois,  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror  
 akouch, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zumestein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580, MUID:98044033, PMID:9384377  
 A:Accession: D69612  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-443 <KUN>  
 A:Cross-references: GB:D29104; GB:AL009126; NID:92632267; PIDN:CAB1786.1; PID:92632277  
 A:Experimental source: strain 168  
 A:Genetics:  
 A:Gene: daaa  
 A:Start codon: TTG  
 C:Superfamily: penicillin-binding protein 5  
 C:Keywords: cell wall synthesis; hydrolase; serine carboxypeptidase  
 F:67/Active site: Ser #status experimental

Query Match 10.3%; Score 92.5; DB 2; Length 443;  
 Best local Similarity 25.0%; Pred. No. 13;  
 Matches 36; Conservative 26; Mismatches 49; Indels 33; Gaps 6;

QY 6 DPA-RNTVKEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEK 64

Db 304 DYA FDNF S M K E I Y A B D ----- Q V K G H K - T I S V D K G K E K E V G I V T N K A F S L P V K N G E - E K 356  
QY 65 GY Q F D G W E I S G F E G K K D A G Y I N L S K D T F I K P V F K K I E E K K E E N K P T F D V S K K K D N P Q V 124  
Db 357 N Y K A K ----- V T L N K D N L T A P V K K G T K V G K ----- L T A E Y T G D E K 391  
QY 125 N H S O L N E S H R K E D L O R E H S O K S D 148  
Db 392 D Y G F L N S D L A G V D L V T K E N V E K A N 415

Search completed: February 10, 2004, 10:58:37  
Job time : 14.5638 secs

***This Page Blank (uspto)***

PRINTS; PRO1228; EGGSHELL.  
 KW Hypothetical protein; Signal; Transmembrane.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT CHAIN 18 385 HYPOTHETICAL GLYCINE-RICH PROTEIN  
 FT TRANSMEM 165 185 POTENTIAL.  
 FT DOMAIN 23 190 GUY-RICH.  
 FT SEQUENCE 385 AA; 36963 MW; 32AB25AE9B5073FE CRC64;

Query Match 11.7%; Score 99.5; DB 5; Length 385;  
 Best Local Similarity 27.7%; Pred. No. 13;  
 Matches 46; Conservative 26; Mismatches 51; Indels 43; Gaps 8;

QY 2 TYKEPLIND-----TGEVSELKPHRYVTITQ--NGKMSSTVSEDPF 44  
 DB 232 TLKEKAGQDIYVVCNEKAPFTAFETDDFCSLQENVHCITLIRINKEVAK--NEEDKK 289  
 QY 45 LPVYKGELEKGYOPDGMELSGPBGKDGAVINLSKOTFIKPFKIEKKEENKPTD 104  
 DB 290 EBPKEKEKEKEKEKEKE--BDEKDE-----EP--KKEEKEKEKEKE--E 330  
 QY 105 VSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTQVATATLUDK 150  
 DB 331 VERKEE-----EKKDEBPKEKEKEKEKEKEDEVEKSEKVBK 371

## RESULT 39

OS1339 PRELIMINARY; PRT; 4524 AA.  
 AC 081339;  
 DT 01-MAR-2003 (TRENBLREL. 23, Created)  
 DT 01-MAR-2003 (TRENBLREL. 23, last sequence update)  
 DT 01-MAR-2003 (TRENBLREL. 23, last annotation update)  
 DE Hypothetical protein.  
 GN PF31325W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 NCBI\_TaxID=36329;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrett A., Brooks K.,  
 RA Mungall K., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Alkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagels K., James K.D., Johnson D., Kethorou A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinovitch B.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sultoni J.E., Craig A., Newbold C., Barrett B.G.  
 RT "Sequence of plasmodium falciparum chromosomes 1, 3-9 and 13."  
 RL EMBL:4195527-531(2002).  
 DR EMBL; AL929354; CAD51629.1; -.  
 KW Hypothetical protein.  
 SO SEQUENCE 4524 AA; 525287 MW; 542206173C29567A CRC64;

Query Match 11.7%; Score 99.5; DB 5; Length 4524;  
 Best Local Similarity 24.3%; Pred. No. 2e+02;  
 Matches 44; Conservative 27; Mismatches 81; Indels 29; Gaps 7;

QY 1 TTVEKPIIKND-----TGEVSELKPHRYVTITQ--NGKMSSTVSEDPF 43

DB 3812 TNESEVITKDKIIMNDVPTQEBFDIKPEVESNLIESNNNNNNKKSKNNIYKORSM 3871  
 QY 44 ILPVYKGELEKGYOPDGMELSGPBGKDGAVINLSKOTFIKPFKIEKKEENKPTD 103  
 DB 3872 VHYINQV---YFPNSIRYSNFNLSIDLN-MINLANKN-SKITTDLILSKCIHNSMS 3926  
 QY 104 DVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTQVATATLUDKNNISSKSTNNPN 163  
 DB 3927 SIIRKYNMKIKQSVIKKISMKNIIOKET-EKKQDVYKEV-----KNNLRKSDANKN 3979  
 QY 164 K 164  
 DB 3980 K 3980

## RESULT 40

P91488 PRELIMINARY; PRT; 211 AA.  
 AC P91488;  
 DT 01-MAY-1997 (TRENBLREL. 03, Created)  
 DT 01-MAY-1997 (TRENBLREL. 03, last sequence update)  
 DT 01-MAR-2003 (TRENBLREL. 23, last annotation update)  
 DE Hypothetical 24.0 kDa protein.  
 GN T23B3.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscot N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium."  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscot N2;  
 RA Magg L., Le T.;  
 RT "The sequence of C. elegans cosmid T23B3."  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscot N2;  
 RA Waterston R.;  
 RT "Direct Submission."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U08309; AAB42334.1; -.  
 DR WormPep; T23B3.5; C014016.  
 KW Hypothetical protein.  
 SO SEQUENCE 211 AA; 23955 MW; 11B16164A87E5928 CRC64;

Query Match 11.7%; Score 99; DB 5; Length 211;  
 Best Local Similarity 29.4%; Pred. No. 7.6;  
 Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

QY 67 EGKQAGVYINLSKOTFIKPFKIEKKEENKPTDVKSKKNPQVNH-----SQLNES 122  
 DB 69 EGKQAGVYINLSKOTFIKPFKIEKKEENKPTDVKSKKNPQVNH-----SQLNES 122  
 QY 123 HRKEDLQREHSQKSDSTQVATATLUDKNNISSKSTNNPNK 164  
 DB 129 EKKQDKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEK 170

Search completed: February 10, 2004, 10:57:09  
 Job time : 31.0366 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 30.8071 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897  
Sequence: 1 KIVKQDFPARNITVKEFILNK.....ATVLDKNNISSKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL.23.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	2119	2 Q9AHT5	Q9AHT5 streptococc
2	897	100.0	2140	16 Q97RY6	Q97RY6 streptococc
3	894	99.7	2144	16 Q8DOP7	Q8DOP7 streptococc
4	893	99.6	2144	2 Q9S4M8	Q9S4M8 streptococc
5	119	13.3	775	16 Q8CPK8	Q8CPK8 staphylococ
6	114.5	12.8	1038	13 Q90784	Q90784 gallus galli
7	114	12.7	609	5 Q812K8	Q812K8 plasmodium
8	112.5	12.5	346	5 Q9U0G0	Q9U0G0 plasmodium
9	112.5	12.5	354	5 Q25995	Q25995 plasmodium
10	112.5	12.5	354	5 Q81U55	Q81U55 plasmodium
11	112.5	12.5	361	5 Q95P15	Q95P15 plasmodium
12	112.5	12.5	379	5 Q9U6C4	Q9U6C4 plasmodium
13	112	12.5	829	5 Q815F3	Q815F3 plasmodium
14	111.5	12.4	379	5 Q25706	Q25706 plasmodium
15	110.5	12.3	379	5 Q25705	Q25705 plasmodium
16	110	12.3	380	5 Q26019	Q26019 plasmodium

17	109.5	12.2	3008	5 Q81436	Q81436 plasmodium
18	109	12.2	951	5 Q96229	Q96229 plasmodium
19	108.5	12.1	600	5 Q77355	Q77355 plasmodium
20	107.5	12.0	470	10 Q9FUK9	Q9FUK9 arabidopsis
21	107.5	12.0	948	5 Q9U4U6	Q9U4U6 plasmodium
22	106	11.8	3026	5 Q81LS9	Q81LS9 plasmodium
23	105.5	11.8	329	5 Q9NFV9	Q9NFV9 plasmodium
24	105.5	11.8	2081	10 Q91498	Q91498 arabidopsis
25	105	11.7	3127	5 Q81DA0	Q81DA0 plasmodium
26	104	11.6	540	10 Q94C59	Q94C59 arabidopsis
27	104	11.6	540	10 Q9S8A4	Q9S8A4 arabidopsis
28	103.5	11.5	328	5 Q44016	Q44016 dictyosteli
29	103	11.5	235	5 Q81226	Q81226 plasmodium
30	102.5	11.4	385	5 Q93424	Q93424 caenorhabdi
31	102.5	11.4	449	5 Q81HW3	Q81HW3 plasmodium
32	102	11.4	622	16 Q98Q41	Q98Q41 mycoplasma
33	102	11.4	1419	5 Q9UBD4	Q9UBD4 plasmodium
34	102	11.4	1612	5 Q81FNO	Q81FNO plasmodium
35	101.5	11.3	312	16 Q9PPL5	Q9PPL5 campylobact
36	101.5	11.3	382	5 Q9V710	Q9V710 drosophila
37	101.5	11.3	535	5 Q17595	Q17595 caenorhabdi
38	101.5	11.3	556	5 Q9V719	Q9V719 drosophila
39	101.5	11.3	785	5 Q9G082	Q9G082 drosophila
40	101.5	11.3	1377	5 Q81586	Q81586 plasmodium
41	101.5	11.3	1434	5 Q81J77	Q81J77 plasmodium
42	101	11.3	700	3 Q07457	Q07457 saccharomyc
43	101	11.2	1490	5 Q81416	Q81416 plasmodium
44	100.5	11.2	354	2 Q93CL2	Q93CL2 mycoplasma
45	100.5	11.2	1373	5 Q81353	Q81353 plasmodium

## ALIGNMENTS

### RESULT 1

Q9AHT5 PRELIMINARY: PRT: 2119 AA.

AC Q9AHT5;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Serine protease (Fragment).  
GN PRTA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
RA Langemann S., Johnson S., Koenig S.;  
RT "Use of a Whole Genome Approach to Identify Vaccine Molecules  
Affording Protection against Streptococcus pneumoniae Infection.";  
RT Infect. Immun. 69:1593-1598(2001).  
-- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
-- AN AMIDE BOND (BY SIMILARITY).  
CC EMBL: AF291699; AAK19159.1; --  
DR HSRP; P00782; ZSRT.  
DR MEROPS; S08.064; --  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRPFMs; TIGR01167; LPXTG\_anchor; 1.

DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Cell wall; Peptidoglycan-anchor; Protease.  
FT NON TER 1  
SQ SEQUENCE 2119 AA; 238226 MW; 517F9B7F6B960A6A CRC64;  
  
Query Match 100.0%; Score 897; DB 2; Length 2119;  
Best Local Similarity 100.0%; Pred. No. 7.7e-54;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KIVVDFPANTYKKEFLINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILLPVYKG 60  
DB 1912 KIVVDFPANTYKKEFLINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILLPVYKG 1971  
  
QY 61 ELEKGYQPDGWEISGEGKKDAGYVNLNLSKDTFIRPVFKKIEKKKEENKPTFDVSKKD 120  
DB 1972 ELEKGYQPDGWEISGEGKKDAGYVNLNLSKDTFIRPVFKKIEKKKEENKPTFDVSKKD 2031  
  
QY 121 NPQVNSQINSHRKEDLQREHSQKSDSTKQVTAIVLDKNNISSKSTNNPNK 174  
DB 2032 NPQVNSQINSHRKEDLQREHSQKSDSTKQVTAIVLDKNNISSKSTNNPNK 2085  
  
RESULT 2  
Q97RY6 PRELIMINARY; PRT; 2140 AA.  
ID Q97RY6  
AC Q97RY6;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Serine protease, subtilase family.  
GN SP0641.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RA SEQUENCE FROM N.A.  
RC STRAIN=TIGR4;  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D., Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J., Durkin A.S., Gattam M., Kolonay J.F., Nelson W.C., Peterson J.D., Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D., Holtapple E., Khouiri H., Wolf A.M., Utecherback T.R., Hansen C.L., McDonald L.A., Feldblyum T.V., Angluoli S., Dickinson T., Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.; "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae"; Science 293:498-506(2001).  
RT EMBL; AE007373; AAK74791.1; -.  
DR MEROPS; S08\_064; -.  
DR TIGR; SP0641; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00882; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRPFMS; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Protease; Complete proteome.

SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;  
  
Query Match 100.0%; Score 897; DB 16; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 7.8e-54;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KIVVDFPANTYKKEFLINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILLPVYKG 60  
DB 1933 KIVVDFPANTYKKEFLINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILLPVYKG 1992.  
  
QY 61 ELEKGYQPDGWEISGEGKKDAGYVNLNLSKDTFIRPVFKKIEKKKEENKPTFDVSKKD 120  
DB 1993 ELEKGYQPDGWEISGEGKKDAGYVNLNLSKDTFIRPVFKKIEKKKEENKPTFDVSKKD 2052  
  
QY 121 NPQVNSQINSHRKEDLQREHSQKSDSTKQVTAIVLDKNNISSKSTNNPNK 174  
DB 2053 NPQVNSQINSHRKEDLQREHSQKSDSTKQVTAIVLDKNNISSKSTNNPNK 2106  
  
RESULT 3  
Q8DQ7 PRELIMINARY; PRT; 2144 AA.  
ID Q8DQ7  
AC Q8DQ7;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Cell wall-associated serine proteinase Prta (EC 3.4.21.-).  
GN PRta OR SP80561.  
OS Streptococcus pneumoniae (strain ATCC BAA-235 / R6).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=171101;  
RN [1]  
RA SEQUENCE FROM N.A.  
RC MEDLINE=21429245; PubMed=11544234;  
RX Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S., Dehoff B.S., Batrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C., Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E., LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P., McAhren S.M., McHenry M., Mcleaster K., Mundy C.W., Nicae T.I., Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P., Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G., Zook C.A., Baltz R.H., Tskunas S.R., Roestek P.R. Jr., Skatrud P.L., Glass J.I.; "Genome of the bacterium Streptococcus pneumoniae strain R6"; J. Bacteriol. 183:5709-5717(2001).  
RT EMBL; AE008434; AAK9365.1; -.  
DR HydroLase; Complete proteome.  
KW SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;  
  
Query Match 99.7%; Score 894; DB 16; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 1.3e-53;  
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KIVVDFPANTYKKEFLINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILLPVYKG 60  
DB 1937 KIVVDFPANTYKKEFLINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILLPVYKG 1996  
  
QY 61 ELEKGYQPDGWEISGEGKKDAGYVNLNLSKDTFIRPVFKKIEKKKEENKPTFDVSKKD 120  
DB 1997 ELEKGYQPDGWEISGEGKKDAGYVNLNLSKDTFIRPVFKKIEKKKEENKPTFDVSKKD 2056  
  
QY 121 NPQVNSQINSHRKEDLQREHSQKSDSTKQVTAIVLDKNNISSKSTNNPNK 174  
DB 2057 NPQVNSQINSHRKEDLQREHSQKSDSTKQVTAIVLDKNNISSKSTNNPNK 2110  
  
RESULT 4  
Q9S4M8 PRELIMINARY; PRT; 2144 AA.  
ID Q9S4M8  
AC Q9S4M8;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)



DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Cell wall-associated serine proteinase Prta precursor.  
 GN PRTA.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 CC Streptococcus.  
 NCBI\_TaxID=1313;  
 OX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3.B;  
 RA Behe G., ten Thoren E., Bongaeerts R.J.M., Heinz H.-P., Zysk G.,  
 RT "Cloning and sequencing of a novel surface protease of Streptococcus  
 pneumoniae."  
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL. PEPTIDOGLYCAN BY  
 AN AMIDE BOND (BY SIMILARITY).  
 DR EMBL; AF127143; AAD48399.1; -.  
 DR HSSP; P00782; 2SRT.  
 DR MEROPS; S08.064; -.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR006192; LPTG.  
 DR InterPro; IPR003137; PA.  
 DR InterPro; IPR000209; Peptidase\_S8.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF02225; PA\_1.  
 DR Pfam; PF00082; Peptidase\_S8; 2.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRFAMs; TIGR01167; LPTG\_anchor; 1.  
 DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS00840; PA\_1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 KW Cell wall; Peptidoglycan-anchor; Signal.  
 FT SIGNAL 1  
 FT CHAIN 20 2144  
 FT POTENTIAL.  
 FT CELL WALL-ASSOCIATED SERINE PROTEINASE  
 FT PRTA.  
 SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
 Query Match 99.6%; Score 893; DB 2; Length 2144;  
 Best Local Similarity 98.9%; Pred. No. 1.5e-53;  
 Matches 172; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KIVVDFARNTTVEKFIINKDTGEVSELRPHRVTVTIQNGKEMSSITVSEBDPILPVYKG 60  
 DB 1937 KIVVDFARNTTVEKFIINKDTGEVSELRPHRVTVTIQNGKEMSSITVSEBDPILPVYKG 1996  
 QY 61 ELEKGYQDGMWISFEGEKDAGYVINSKDTFICPVFKIIEKKEENKPTFDVSKKD 120  
 DB 1997 ELEKGYQDGMWISFEGEKDAGYVINSKDTFICPVFKIIEKKEENKPTFDVSKKD 2056  
 QY 121 NPQVNSQLNESHKEDLOREHESQKSDTKDVTATVLDKNNISKSTTNPNK 174  
 DB 2057 NPQVNSQLNESHKEDLOREHESQKSDTKDVTATVLDKNNISKSTTNPNK 2110  
 RESULT 5  
 OSCP88 PRELIMINARY; PRT; 775 AA.  
 ID OSCP88  
 AC OSCP88  
 DT 01-MAR-2003 (TReMBLrel. 23, Created)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Penicillin-binding protein 1.  
 GN SE0856  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 NCBI\_TaxID=1282;  
 OX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,

RA Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AE016746; AA004453.1; -.  
 KW Complete Proteome.  
 SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;  
 Query Match 13.3%; Score 119; DB 16; Length 775;  
 Best Local Similarity 27.0%; Pred. No. 2;  
 Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
 QY 20 KDTGEVSELRPHRVTVTIQNGKEMSSITVSEBDPILPVYK-----GELEKGYQFDGM-- 71  
 DB 614 EDSVNAQSLKP-----ITIGNGKQIKQSGKSTKVLPHSKYMLMTDGLTWP-DWTGWTK 668  
 QY 72 -EISGFE-----GKKDAGYVYN-LSKDTFICPVFKIIEKKEENKPTFDVSKKD 117  
 DB 669 EDVLAFFEDLTRIKYSTKNGFVTNQSISKQIIR-----NKDIEVLSAED 715  
 QY 118 KKDNPQVNSQLNESHKEDLOREHESQKSDTKDVTATVLDKNNISKSTTN 171  
 DB 716 TDDDEKTDSDSKSKDQADBDHSNTSSITKN-----DKSNADSKNDSDD 763  
 RESULT 6  
 ID O90784 PRELIMINARY; PRT; 1038 AA.  
 AC O90784  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
 DE Claustrian.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 OX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Cole G.J.;  
 RL Submitted (OCT-1992) to the EMBL/Genbank/DBJ databases.  
 RP SEQUENCE OF 1-451 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=94157526; PubMed=7906711;  
 RA Burg M.A., Cole G.J.;  
 RT "Claustrian, an antiadhesive neural keratan sulfate proteoglycan, is  
 RT structurally related to MAP1B."  
 RL J. Neurobiol. 25:1-32(1994).  
 DR EMBL; X67778; CAA47988.1; -.  
 SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;  
 Query Match 12.8%; Score 114.5; DB 13; Length 1038;  
 Best Local Similarity 23.5%; Pred. No. 5.7;  
 Matches 47; Conservative 29; Mismatches 65; Indels 59; Gaps 6;  
 QY 2 IVVDFARN---TVKEFIINKDTGEVSELRPHRVTVTIQ-----GELEKGYQFDGM-- 39  
 DB 514 VTQKDLTGNINSPAVKQAKLKQRTDSKESLKPAAKTTTKQCKQKRLKKKHWISLQSLVQ 573  
 QY 40 -----GKMSSTVSEBDPILPVYKGELEKGYQFDGMWISGFGK 79  
 DB 574 LEKPKLESKEKTPYKKEKAVKPEKTVIAKDV-----TKKEQLGSEISEKQAS 625  
 QY 80 KDAGVYNLSKDTFICPVFK-KIEKKEENKPTFDVSKKDNPOVNSQLNESHKEDL 138  
 DB 626 EKQDVPKPVTKKSKYKKEVAKPEKDKKPKKKEVSKKEKPLI---KKEKKKKEDI 682  
 QY 139 QRE-----EHSQKSDTKDV 153  
 DB 683 KKEVKEVKEKKEKKEKKEV 702

RESULT 7	
0812K8	PRELIMINARY; PRT; 609 AA.
ID 0812K8	
AC 0812K8	
DT 01-MAR-2003 (TRENBLREL. 23, Created)	
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)	
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)	
DE Large cyclophilin-like protein.	
GN PFI1490C.	
OS Plasmodium falciparum (isolate 3D7).	
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX NCBI_TaxID=36329;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=22255708; PubMed=12368867;	
RA Hall N., Pain A., Berriman M., Churruarin C., Harris B., Harris D.,	
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,	
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,	
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,	
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,	
RA Feltwell T., Goble A., Goodhead I., Gilliam R., Hamlin N., Hance Z.,	
RA Harper D., Hauser H., Hornsby T., Holtroyd S., Horrocks P.,	
RA Humphrey S., Jagals K., Jones K.D., Johnson D., Kethornou A.,	
RA Knights A., Konfortov B., Kyes S., Lake N., Lawson D., Lennard N.,	
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,	
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,	
RA Rajadream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,	
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,	
RA Taylor K., Tiley A., Unwin L., Whitehead S., Woodward J.,	
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;	
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";	
RL Nature 419:527-531(2002).	
RQ EMBL; AL929356; CAD51984.1; --	
SR SEQUENCE 609 AA; 72552 MW; 8CDF68E85FF9A021 CRC64;	

[illegible][illegible]

Q25995	PRELIMINARY;	PRT;	354 AA.
AC	Q25995		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)		
DE	Antigen.		
OS	Plasmodium falciparum (isolate NF54).		
CC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.		
OX	NCBI_TaxID=5843;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NF54;		
RX	MEDLINE=95196774; PubMed=7891748;		
RA	McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,		
RA	Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;		
RT	"Molecular variation in a novel polymorphic antigen associated with		
RT	Plasmodium falciparum merozoites."		
RL	Mol. Biochem. Parasitol. 68:53-67(1994).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NF54;		
RX	MEDLINE=98156743; PubMed=9497029;		
RA	McColl D.J., Anders R.F.;		
RT	"Conservation of structural motifs and antigenic diversity in the		
RT	Plasmodium falciparum merozoite surface protein-3 (MSP-3)."		
RL	Mol. Biochem. Parasitol. 90:21-31(1997).		
DR	EMBL; L28825; AAC09377.1; "		
SQ	SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;		
QY	Query Match	12.5%;	Score 112.5; DB 5; Length 354;
	Best Local Similarity	22.4%;	Pred. 2.5;
	Matches 41; Conservative 35; Mismatches 62; Indels 45; Gaps 7		
QY	10 NTTVKEPLIN-KDNGEVELKPHRYTVTIQNGKMSSTIVSEDFILPYKKELEKGYQF	68	
DB	153 STKKEVEKAKNAVEKAKNAVQKANAQVLAKEAS-----YDTIL-----	194	
QY	69 DGWEISGF--EGKKDAG-----YVINKSDTFFIKFKEKEKE-----E	108	
DB	195 -GWEEGGVPEHKKEENMLSHLYVSSKDKENISKENDVDLDEKKEAEAEETEEBLEEKNE	253	
QY	109 NKPTFVSKKDNFQVNHSSQINESHKEDLQREHSHQKSDSTKYDVTATYVLKNNISSKST	168	
DB	254 EETSESEIDEEDEEEEEEKEENDKKKEQKEQGENENNDDQKKDMEA-----QNLISKQ	308	
QY	169 TNN 171		

DB 309 NNN 311

RESULT 10

Q81J55 PRELIMINARY; PRT; 354 AA.

AC Q81J55; 01-MAR-2003 (TREMblrel. 23, Created)

DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)

DE Merzoite surface protein 3.

GN PF10\_0345.

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxId=36329;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=3D7;

RX MEDLINE=2255705; PubMed=12368864;

RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., Carlton J.M., Patil A., Nelson K.E., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angioli S., Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B., Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Frazer C.W., Barrett B.;

RT "Genome sequence of the human malaria parasite Plasmodium falciparum."

RL Nature 419:498-511 (2002).

DR EMBL; AE014834; AAN35542.1; -

SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match

Best Local Similarity 12.5%; Score 112.5; DB 5; Length 354; Matches 41; Conservative 35; Mismatches 62; Indels 45; Gaps 7;

QY 10 NTTVKFEFLIN-KDTGEVSLKPHRYVTITQNGKMSSTIVSEDFILPYKGELEKGYOF 68

DB 153 STKTEYKAKNAAYEKANAYOKANQAVLAKAKSS-----YDYLL----- 194

QY 69 DGMETISGF--EGKKDAG-----YVINLSKDTFIKPYVKIEKKEE-----E 108

DB 195 -GWERGGGVPEHKKEENMLSHLYSSKDKENISKENDVDLDEKEBAETEEBELKNE 253

QY 109 NKPTDVSKKQDNPOVNHSQLNESHKEDLQREHSHQSKSDTKDVTATVLDKNNISSKST 168

DB 254 EETEESEISDEDEEEKEEENDKKEQKEQSENNDQKKDMEA-----QNLISKNG 308

QY 169 TNN 171

DB 309 NNN 311

RESULT 11

Q95P15 PRELIMINARY; PRT; 361 AA.

AC Q95P15; 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)

DE Merzoite surface protein 3 (Fragment).

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxId=5833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=FVO;

RA Hiseida H., Saul A., Long C.A., Miller L.H., Stowers A.W.;

RT "Merzoite Surface Protein 3 and Protection Against Malaria in Actus Monkeys."

RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL; AY044180; AAK94780.1; -

FT NON TER 1

FT NON TER 361

SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match

Best Local Similarity 12.5%; Score 112.5; DB 5; Length 361; Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8;

QY 21 DTGEVSELKPHRYVTITQNGKMSSTIVSEDFILPYKGELEKGYO----- 67

DB 147 ETGE-----RNSRNNFYTTYKX-----YAGKVKDYERAKNAYOKANQAV 187

QY 68 -----FD-----GMEISGF--EGKKDAG-----YVINLSKDTFIKPYVKIEKKEEN 109

DB 188 LKAKEASDYITLIGWFGGGVPEHKKEENMLSHLYSSKDKENISKENDVDLDE-KEBA 246

QY 110 KPTFDVSKKQDNPOVNHSQLNESHKEDLQREHSHQSKSDTKDVTATVLDKNNISSKST 161

DB 247 EETEESELEKNEETEESEISDEDEEEKEEENDKKEQKEQSENNDQKKDME 306

QY 162 --NISKSTNN 171

DB 307 AQNLISKNNNN 318

RESULT 12

Q9U6C4 PRELIMINARY; PRT; 379 AA.

AC Q9U6C4; 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DE Polymorphic antigen.

GN MSP-3.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxId=5833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PCCL/HN;

RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;

RT "Sequence of Plasmodium falciparum secreted polymorphic antigen gene."

RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF188190; AAF04099.1; -

SQ SEQUENCE 379 AA; 43316 MW; C152A54B1F9D5F25 CRC64;

Query Match

Best Local Similarity 12.5%; Score 112.5; DB 5; Length 379; Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;

QY 29 KPHRYVTITQNGKMSSTIVSEDF-----ILPYKGELEKGYOFD-GMEISGF- 76

DB 171 KPSRLNLSRKTKETAEAY--EKDYERAKNAYOKANQAVLAKAKSSDYITLIGWFGGV 228

QY 77 -EGKKDAG-----YVINLSKDTFIKPYVKIEKKEENKPTFDVSKKQDNPOVNHSQLN 130

DB 229 PEHKKEENMLSHLYSSKDKENISKENDVDLDE-KEBAETEEBELKNEETEESEIS 287

QY 131 -----ESHKEDLQREHSHQSKSDTKDVTATVLDKNNISSKSTNN 171

DB 288 EDEEESELEKNEETEESEISDEDEEEKEEENDKKEQKEQSENNDQKKDMEA-----QNLISKNNN 336

RESULT 13

Q81SF3 PRELIMINARY; PRT; 829 AA.

AC Q81SF3; 01-MAR-2003 (TREMblrel. 23, Created)

DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)

DE Hypothetical protein.

Matches	41;	Conservative	35;	Mismatches	59;	Indels	39;	Gaps	8
QY 29	KPHRVTVTIQNGKEMSTIVSEEDF-----ILPVKGELKGTQFD-GWEISGF	76							

RT "Molecular variation in a novel polymorphic antigen associated with  
RT *Plasmodium falciparum* merozoites.";  
RT

RL M01. Biochem. Parasitol. 68:53-67(1994).  
 RN [2] SEQUENCE FROM N.A.  
 RC STRAIN=FC27;  
 RX MEDLINE=98156743; PubMed=9497029;  
 RA McCall D.J., Anders R.F.;  
 RT "Conservation of structural motifs and antigenic diversity in the  
 RL Plasmodium falciparum merozoite surface protein-3 (MSP-3)."  
 DR M01. Biochem. Parasitol. 90:21-31(1997).  
 KW Signal.  
 FT CHAIN 1 25 POTENTIAL.  
 SQ SEQUENCE 360 AA; 43290 MW; 0986CA1393094CA2 CRC64;  
 Query Match 12.3%; Score 110; DB 5; Length 380;  
 Best Local Similarity 24.2%; Pred. No. 4;  
 Matches 45; Conservative 34; Mismatches 59; Indels 48; Gaps 9;  
 QY 7 PARNTTKEFLIN-KOTGEVSELPKRYVTITQNGKMSSTIVSEDPILPYKGELEKG 65  
 DB 179 FSRRT--KEYAEKAKNAVEKAKNAYOKANQAVLKAKKAS-----YDYIL----- 221  
 QY 66 YQFQSGEISGF--EGKXGAG-----YVINLSKDTFKPKVFKKIEKKKEENKPTDVSCK 118  
 DB 222 ---GMEFGGVPEHKKEENMLSHLYSSKDKENSKENDVLDLDE--KESEAEETEEL 276  
 QY 119 KDNPOVNHNSLN-----ESHKEDLQREHSOKSDSTKYDVAIVLDKNNTSS 165  
 DB 277 EKNEETSEISEDEEEBEKEEENKKEKEKEQENENNDDQKKDMEA-----QNLIS 331  
 QY 166 KSTTN 171  
 DB 332 KNQNN 337  
 RESULT 17  
 ID Q81436 PRELIMINARY; PRT; 3008 AA.  
 AC Q81436;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PF60325W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 CX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrett B.;  
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22255708; PubMed=12368867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungal K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Felwell T., Goble A., Goodhead I., Gilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagsels K., James K.D., Johnson D., Kerhorn A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lemard N.,  
 RA Line A., Maddison M., McLean J., Mooney P., Mouton S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sultson J.E., Craig A., Barrett B.G.;  
 RL "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
 RT Nature 419:527-531(2002).

DR EMBL; AL929351; CAD51431.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 3008 AA; 356023 MW; 60BCBBE15C599B4 CRC64;  
 Query Match 12.2%; Score 109.5; DB 5; Length 3008;  
 Best Local Similarity 24.6%; Pred. No. 39;  
 Matches 52; Conservative 37; Mismatches 71; Indels 51; Gaps 10;  
 QY 1 KIVKDPARNTTKEFLINKTGEVSELK-----PRVVTITQNGKMSSTIVSE 50  
 DB 2218 KVLIEKCEKESLISHSTEND-----ETKDDNNNNNNNNKHYGYNNIDHKNINKQEVTT 2272  
 QY 51 EDF-----LLPYKGELEKGYQPDGMEISGFEKGDAGYVNLK-----DTFKP 96  
 DB 2273 QDANSSNPNLNNITQKKKRTNSKYIFCEIQVRE-----YDIEKLEKFGASIGP 2326  
 QY 97 VFKKIEKKEEENKPTDVSCKKNPNQVNSQ---LNEHREKEDLQREH---SOKSDS 149  
 DB 2327 VFTD--EENKKEENK--EVNKEENKKEENKKEENKKEENKKEENKKEENK 2383  
 QY 150 TKDYTAIVLDKNNTSK-----STNNPNK 174  
 DB 2384 KKEVKNKEENKKEENKKEENKKEENKKEENK 2414  
 RESULT 18  
 ID Q96229 PRELIMINARY; PRT; 951 AA.  
 AC Q96229;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PF60680W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 CX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=99021743; PubMed=9804551;  
 RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,  
 RA Koonin E.V., Shalim S., Mason T., Yu K., Fujii C., Pederson J.,  
 RA Shen K., Jing J., Aston C., Lai Z., Schwartz R.C., Petrea M.,  
 RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
 RA Frazer C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
 RT falciparum.";  
 RL Science 282:1126-1132(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eissen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shalim S.J., Suh B., Peterson J., Anguioni S.,  
 RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Frazer C.M., Barrett B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum.";  
 RL Nature 419:498-511(2002).  
 DR EMBL; AE001410; AAC71925.2; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 951 AA; 112486 MW; AC8D889358A84F4F CRC64;  
 Query Match 12.2%; Score 109; DB 5; Length 951;  
 Best Local Similarity 26.1%; Pred. No. 12;  
 Matches 49; Conservative 32; Mismatches 49; Indels 58; Gaps 11;

QY 5 KDFARTVTKFPIKNGTGEVSELKPHRYT-VTIQNGKEMSTIVSEDFILPYVKGLE 63  
 DB 115 KODNNNNNGTKOIEKNGKINKSUL-HRONELNLSGK-----NEQOI-----NKKE 159  
 QY 64 KGYQDFGWEI-SGFEGKQDAGVNLKSDTFIKPVFKKIEEKE-----ENKP 111  
 DB 160 KGMQ-----DISNNAENKQD-----VKEGVKEIEEKKKEKISDDHKVYENK 203  
 QY 112 TPD---VSKKQDPPVNHNSQALNESHKRDQOR-EHSHQKSDSTKVTATVLDKNNISK 166  
 DB 204 SDHKVYENKKSDDHKVYENKKSDDHKIEVKVKEEHEDEE-----DKKKEKE 254  
 QY 167 STNNPNK 174  
 DB 255 NKKKDKNK 262

RESULT 19  
 ID 077355 PRELIMINARY; PRT; 600 AA.  
 AC 077355;  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 GN PFC0465C, MAL34.20.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=99376085; PubMed=10448855;  
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Fellwell T.,  
 RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
 RA Horrocks P., Jagsels K., Jassal B., Kyes S., McLean J., Moule S.,  
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,  
 RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,  
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;  
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium  
 RT falciparum."  
 RL Nature 400:532-538(1999).  
 DR EMBL: AL008970; CA15610.2; -  
 DR InterPro: IPR002483; PFI.  
 DR Pfam: PF01480; PFI; 1.  
 DR SMART: SM00311; PFI; 1.  
 KW Hypochemical protein.  
 SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;

Query Match 12.1%; Score 108.5; DB 5; Length 600;  
 Best Local Similarity 29.3%; Pred. No. 8.3;  
 Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;

QY 47 IYSEEDFILPVY-----KGELEKGYQDFGWEISGFEGK-----DAGVINLSGDTFIKPY 97  
 DB 60 ILFEPEDILVEYICISQKQSKKEK--DGEEDYILNAKKLKTINLTGFTGKKSDFIIEEL 116  
 QY 98 FKXI--EKKKE-----ENKPTFVSK-KQDPPVNHNSQALNE-----SHRK 135  
 DB 117 LELLINKEKKEHIAATLNENK-TNDIKYKNNENINENVENNDISNKKDEHSHQK 175  
 QY 136 E-----DLQREH-----SQKSDSTK---DVTATVLDKNNISSTTN 170  
 DB 176 EHNINNVNLKKEKEYTDIQDKKKKKSLSQKSDSYKRPFNKRTKISIER-SLSNRYDE 234  
 QY 171 NPNK 174  
 DB 235 KTNK 238

RESULT 20  
 Q9FJK9

ID Q9FJK9 PRELIMINARY; PRT; 470 AA.  
 AC Q9FJK9;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE GB|AA20218.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustoid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=99087489; PubMed=9872454;  
 RA Nakamura Y., Sato S., Asamizu E., Kaneo T., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
 RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
 RT physically assigned pl and TAC clones."  
 RL DNA Res. 5:297-308(1998).  
 DR EMBL: AB015468; BAB10694.1; -  
 SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72E35AC54 CRC64;

Query Match 12.0%; Score 107.5; DB 10; Length 470;  
 Best Local Similarity 20.1%; Pred. No. 7.5;  
 Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 29 KPHRYVTITQNGKEMSTIVSEDFILPYVKGLEKGYQDFGWEISGFE-----GKK 80  
 DB 82 RENRYVTITQNNNSNGSK-----YVDDLARIIRYDE-EATGSGSAGRIDHPNQK 129  
 QY 81 DAGVINLSKQFIKPVFKKIEEKEENKPTFVSKKKN----- 121  
 DB 130 NGITEKAPENPIETSHRVNDNRINNQNPTAKSENASVRSRGADHKRAEWCK 189  
 QY 122 PNVNHSQALNE-----SHRKEDLOREHSHQKSDSTKYATVLDKNNISSTTNPNK 174  
 DB 190 PVNENDQVQTSAAKSHKKNVTSKPKRQDEGVKTEAKQKDNKKEKKEKTSINK 248

RESULT 21  
 ID Q9U4U6 PRELIMINARY; PRT; 948 AA.  
 AC Q9U4U6;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Ornithine decarboxylase.  
 GN ODC.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FCUPL/RSA;  
 RA Birkoletz L., Joubert F., Neitz A.W.H., Louw A.I.;  
 RT "Molecular characterization of Plasmodium falciparum ornithine  
 RT decarboxylase cDNA obtained by RACE."  
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF139900; AAF14518.1; -  
 DR HSSP: P07805; 1F3T.  
 DR InterPro: IPR000183; Decarboxylase2.  
 DR Pfam: PF02784; Orn\_Arg\_dec\_N; 1.  
 DR Pfam: PF00278; Orn\_DAP\_Arg\_dec; 1.  
 DR PRINTS: PR01179; ODACRBLASE.  
 SQ SEQUENCE 948 AA; 110350 MW; 43P103DB83F12835 CRC64;

Query Match 12.0%; Score 107.5; DB 5; Length 948;  
 Best Local Similarity 22.0%; Pred. No. 16;  
 Matches 54; Conservative 32; Mismatches 71; Indels 89; Gaps 11;

QY 4 VKDFARTVTKFPIKNGTGEVSEL-----KPHRYVTITQNGKEMSTIVSEDFI 54

```

Db 547 VEDMSNNMGFNYYIINLGGGYDEELEYDNKAKKDKHVCYTLISLQIKDKIQFLNEEFTL 606
Qy 55 LPVY-----KGELEKGYQFDGWEISGPEGKD-----AGYINLSKOTFIPVPEKK 100
Db 607 KTKYGYSEFEKISLAINNSIDHY-----FSHMKDNRVICEPGRYVAASSTLAVXIGRR 662
Qy 101 IEKKEEENKPTF-----DVSKKDNPOVNSQLNESHREKED-----137
Db 663 -----RPTFGIMLKDLKAHYDPLNFAQGEKKQDEPKINNNNDNNNDNNNDNN 713
Qy 138 LQREHSQK-----SDST-----KDVATVLDK--NNIS-SKS 167
Db 714 NNNNNNNQGGGGINMDLIITSTNDSTYKKNDSHSSGYIQVGSCTIRKEDDNKIMTH 773
Qy 168 TTNPN 173
Db 774 TNNPN 779

```

## RESULT 22

```

Q8ILS9 PRELIMINARY; PRT; 3026 AA.
AC Q8ILS9;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN Pf14_0165.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung B., White O., Berriaman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan W.-S., Nene V., Shalim S.J., Suh B., Peterson J., Angiuoli S.,
RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Frazer C.M., Barrett B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014818; AAN36777.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3026 AA; 357633 MW; 9ECABD915C3C25CE CRC64;

```

```

Query Match 11.8%; Score 106; DB 5; Length 3026;
Best Local Similarity 22.7%; Pred. No. 69;
Matches 44; Conservative 34; Mismatches 86; Indels 30; Gaps 8;

```

```

Qy 1 KIVKDFPARNNT-----VKEFIINKDTGEVSELKPHRVTVYTIQNGKMSSTIVSEDF 53
Db 146 KVLIKESKDLITQDYNLKDIKETEKRTMKSSKYQYNNMNCSTRKSSFNVTDN---202
Qy 54 ILPVYKGELEKGYQFDGWEISGPEGKDAGYINLSKOTFIPVPEKKIEKKEEENKPTF 113
Db 203 ---ICHGN-EKYKMSDNKQICIEIKKQOLIIDEIC--TMVKNANKKIKNOVEEYKNRV 256
Qy 114 DVSKKED---NPOVNSQLNESHREKEDLQREHSQKS--DSTKDVATVLDKNNISKST 168
Db 257 SVINKDKNTIQNSDINNTO-NILHNEDEIBEYKANENDIHTVKTITKEYSSNSGSSSD 315
Qy 169 T-----NNPNK 174
Db 316 TTLSYSEYNNKKNK 329

```

```

RESULT 23
Q9NFV9 PRELIMINARY; PRT; 329 AA.
AC Q9NFV9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
GN MSP3.
OS Plasmodium falciparum (isolate 7G8).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=57266;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7G8;
RX MEDLINE=20416497; PubMed=10960178;
RA Okenu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in
RT Plasmodium reichenowi and Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252287; CAB85901.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 329 AA; 36916 MW; CSB045DB5E21A159 CRC64;

```

```

Query Match 11.8%; Score 105.5; DB 5; Length 329;
Best Local Similarity 24.2%; Pred. No. 7;
Matches 43; Conservative 31; Mismatches 63; Indels 41; Gaps 8;

```

```

Qy 10 NTTVKEFLIN-KDTGEVSELKPHRVTVYTIQNGKMSSTIVSEDFILPVYKGELEKGYQF 68
Db 143 STKTEKVEAKKKNAYEEKKNAYOKANQAVLAKAKASS-----YYIL-----184
Qy 69 DGWEISGF--EGKKDAG-----YVINLSKOTFIPVPEKKIEKKEEENKPTFVSKKON 121
Db 185 -GWEFGGVGPVPHKKEENMLSHLYSSKOKENISKENDVLDE--KEEEAEFEETEELBEKN 242
Qy 122 POWNHSQLNESHREKEDLQREH-----SQKSDSTQDVATVLDKNNISKSTTN 171
Db 243 EETEESEISEDEEEBEKEEKEQAKEQSENNNDQKKDMA-----QNLISKQONNN 295

```

## RESULT 24

```

Q9LH98 PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl.
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL; AP002057; BAB03174.1; -.
SQ SEQUENCE 2081 AA; 232851 MW; D3603E1F85EFFF29 CRC64;

```

```

Query Match 11.8%; Score 105.5; DB 10; Length 2081;

```

Best Local Similarity 25.4%; Pred. No. 50;  
Matches 48; Conservative 31; Mismatches 71; Indels 39; Gaps 8;

QY 2 IYVKDFARNTTKEFLINKDTGEVSELKPHRTVTIIONGKMSSTIVSEDFILPYKGB 61  
DB 1639 IYVRNGKKEKESIK-----GSEDK-----TVEINGSELSTEGSKD-----GK 1678  
QY 62 LKGVQFDGWEISGEGSKD-----AGVILNSDITFKIPVKIKIEKKERENK 111  
DB 1679 IEKGR--EKKENSTKSGSKDKIEEGEKKENSTKSSKDGKINEIHGDKKATMEGSKD 1736  
QY 112 --TFDYSK--KDDPQVNHSQLNESHKREDQ-----REHSOKSDSTKDYATVLDKNIS 164  
DB 1737 GGINSGKSKDSKSKSVINQVNDLSKDSKNDINEINNGKDSYKDNTEIQGDNLS 1796  
QY 165 SKSTNNPN 173  
DB 1797 TNSTSEPN 1805

## RESULT 25

Q8IDA0 PRELIMINARY; PRT; 3127 AA.  
AC Q8IDA0;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, last sequence update)  
DE 01-MAR-2003 (Tremblrel. 23, last annotation update)  
DE Guanylyl cyclase (EC 4.6.1.2).  
GN GC-BETA.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RA SEQUENCE FROM N.A.  
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,  
RA Bertrman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
RA Omond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,  
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL844509; CAD52725.1; -.  
DR Lyase.  
SQ SEQUENCE 3127 AA; 370597 MW; F0375C72B9007560 CRC64;

Query Match 11.7%; Score 105; DB 5; Length 3127;  
Best Local Similarity 23.7%; Pred. No. 84;  
Matches 55; Conservative 36; Mismatches 73; Indels 68; Gaps 10;

QY 9 RNTTVK--EFLINKDTGEVSELK-----HRYVTIIONG-----KMSSTIVSE 50  
DB 1950 RNTTKKSDYLLDNGESKFKRNTSYVLESPLHLIGDIVNNIKRKKKKEIKTIVSD 2009  
QY 51 EDLILPYKGELEKGYQFDGWEISG--FEKKDAGVINLSKDTFIKPVFKIEKKEE 107  
DB 2010 DMFTSPVNIKEYNNEQERKEIVGNLSYDKTKKIPPIFKTEGRKK--KKIEKKKK 2067  
QY 108 ENK-----PTF-----DVSKKDNQVNHSQL----- 129  
DB 2068 EKKENNNFLYNDYSSPYKGNENNFIYTIERRKDFOKKFDHPNPFKFLHNTY 2127  
QY 130 ---NESHK---EDLQREHSOKSDSTKD-VTATVLDKNISKSTNNENK 174  
DB 2128 PMNKKNKKNKKNKKNKVRNEXYPTNTSSKDGVSYNPLSDLSFSSDNEYSUNE 2179

## RESULT 26

Q94C59 PRELIMINARY; PRT; 540 AA.  
AC Q94C59;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
DE Putative phosphatidylinositol-transfer protein.  
GN T518.14.  
OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI\_TaxID=3702;  
RN [1]  
RA SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,  
RA Bowers L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,  
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,  
RA Lam B., Lin Y., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,  
RA Palm C.J., Sakurai T., Satou M., Seki M., Shim P., Southwick A.,  
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
RA "Full length cDNA of gene T518.14 (GI:4587525)".  
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AY035162; AKS59666.1; -.  
DR InterPro; IPR001251; CRAL\_TRIO.  
DR InterPro; IPR001071; RecBind/loctrans.  
DR Pfam; PF00650; CRAL\_TRIO; 1.  
DR Pfam; PF03765; CRAL\_TRIO\_N; 1.  
DR PRINTS; PR00180; CRETINALDHP.  
DR SMART; SM00516; SEC14; 1.  
DR PROSITE; PS50191; CRAL\_TRIO; 1.  
SQ SEQUENCE 540 AA; 61166 MW; 0C2592EB18ACFAC6 CRC64;

Query Match 11.6%; Score 104; DB 10; Length 540;  
Best Local Similarity 25.3%; Pred. No. 15;  
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 11 TVKEFLINKDTGEVSELKPHRTVTIIONGKMSSTIVSEDFILPYKGELEKGYQFDG 70  
DB 27 TVTKA-VVEETVVEDESKP-----EGVEKSAFKEESDFADLKESEKK----- 70  
QY 71 WEISGFEKKDAGVINLSKDTFIKPVFK--IEKKERENKPFDSYKKNQVNH 127  
DB 71 -LMSDKSLTEBAIVDN-----TLTKTKKSSPMKEEYVKAPEAVEKKKE--EAAEE 123  
QY 128 QLNESHKREDLQREHSOKSDSTKDYATVLDKNISS 165  
DB 124 KVEEKKSEAVTEEPKAEVTEAVTEETIIRKEEYTT 161

## RESULT 27

Q9SA84 PRELIMINARY; PRT; 540 AA.

AC Q9SA84;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
DE T518.14 protein (Hypothetical protein).  
GN T518.14 OR ATG30650.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI\_TaxID=3702;  
RN [1]  
RA SEQUENCE FROM N.A.  
RA STRAIN=CV. Columbia.  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Li J., Kremesnekaia I., Iuros J., Ngan I., Gonzalez A., Altafi H.,  
RA Araujo R., Chao O., Conn L., Conway A.B., Dunn P., Hansen N.,  
RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,  
RA Davis R.W., Ecker J.R., Federpie N.A., Theologis A.;  
RA "Arabidopsis thaliana chromosome 1 BAC T518 sequence".  
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RL [2]

RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan Y.W., Lee J.M.,  
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,  
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,  
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,  
RA Ecker J.R., Theologis A.;





```

RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Thomas K.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP REVISIONS.
RC STRAIN=BRISTOL N2;
RA Jones S.J.M.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 281053; CAB02877.1; -.
DR HSSP; P10968; 2COW.
DR WormRep; E02A10.2; CE09116.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHELL.
KW Hypothetical protein; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 385
FT TRANSMEM 165 185
FT DOMAIN 23 190
FT SEQUENCE 385 AA; 36963 MW; 32AB25AE9B5073FE CRC64;
SQ

Query Match 11.4%; Score 102.5; DB 5; Length 385;
Best Local Similarity 26.2%; Pred. No. 13;
Matches 48; Conservative 31; Mismatches 55; Indels 49; Gaps 9;

QY 1 KIVKDPFANTYK-----TYKEFLTKND-----TGEVSELEKPRVTVTIO- 38
DB 215 RILNNTKTKVAESIKTLKEKVAGDYIVVCKNEKAPFAETADPCSLQKENVHCTIIR 274
QY 39 -NKEKMSSTIVSEDFILPYKGELEKGYQFDGMEISGFEKKDAGVYINLSKDTFLKPY 97
DB 275 INKEVAKR--NEEDKKEEPKKEEKEKEVEKKE--EDEKDE-----EP- 316
QY 98 FKIEEKKEREENPTDVSKKKNDPOVNSQLNESHRKEDLQREHSQKSDSTKDYATAY 157
DB 317 -KEEKEKEKEKE--EVEKKE-----BEKDEPPEKKEKEKEBEKEDVEKSEKV 368
QY 158 LDK 160
DB 369 EEK 371

RESULT 31
Q81HW3 PRELIMINARY; PRT; 449 AA.
AC O81HW3;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN Pfl1_0413.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=36329;
RX MEDLINE=22255705; Pubmed=12368864;
RC STRAIN=3D7;
RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyte S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angioli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings I.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Frazer C.M., Barrett B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum";
RL Nature 419:498-511(2002).
DR EMBL; AB014842; AAN3596.1; -.

```

```

KW Hypothetical protein.
SQ SEQUENCE 449 AA; 52812 MW; 85B62272D6257C68 CRC64;

Query Match 11.4%; Score 102.5; DB 5; Length 449;
Best Local Similarity 24.9%; Pred. No. 16;
Matches 52; Conservative 30; Mismatches 58; Indels 69; Gaps 9;

QY 1 KIVKDPFANTYK--FILNKDTGSEVSELKPRVTVTIOQKMSSTIVSEDFILPY 58
DB 31 KSVRVDVADFLNVESESYSENK-----KELLTKIKER--ILILY 70
QY 59 -----KGELEKGYQFDGMEISGFEKKDAGVYINLSKDTFLKPYFKIF- 102
DB 71 NEQONEEKKKKKKNFQSD--ISEKKKSKADLQKSESNNMLKKRKIISEEBETINIS 127
QY 103 --EKKEE-----NKPTDY-----SKKNDPOVNSQLNESHRKEDLQREH 143
DB 128 DDEEEEEYORKKQKQKNSNVSTLSLEKKKKRRDSSENNNDYNEIEDDEQEEH 187
QY 144 SOKSDTKDVTATVLDKNNISSKSTTNP 172
DB 188 EEES-----LSKKSKKKNSTSP 207

RESULT 32
Q98QAL PRELIMINARY; PRT; 622 AA.
AC Q98QAL;
DT 01-OCT-2001 (T-EMBLrel. 18, Created)
DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE LIPOPROTEIN.
GN MYPV 4650.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
ON NCBI_TaxID=2107;
RX MEDLINE=21267165; Pubmed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Gallison F.,
RA Moszer I., Dydwig K., Wroblewski H., Viart A., Kocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
Mycoplasma pulmonis";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL; AL445564; CAC13638.1; -.
KW Complete proteome.
SQ SEQUENCE 622 AA; 73762 MW; 42BD8930861960D CRC64;

Query Match 11.4%; Score 102; DB 16; Length 622;
Best Local Similarity 24.9%; Pred. No. 24;
Matches 48; Conservative 35; Mismatches 66; Indels 44; Gaps 11;

QY 9 RNTYKVFIL-NKDTGSEVSEL-----KPRV--TVTIONG--KEM--SPT--V 48
DB 26 KDSAKNFVFKSSDLISISEKQFKKNLDDKQKIIEGVTFINSETIEYIKKETIILN 85
QY 49 SEEDFILPYKGELEKGYQFDGMEISGFEKKDAGVYINLSKDTFLKPYFKIEKKEE 108
DB 86 SEDNIIFSLNLENNAKFQLDPE-----VSDKEFKIKFQGIINSQTBO 129
QY 109 NKPTDVSKKKD--NPQVNSQLNES--HKEDLQREHSQKSDSTKDYATVLDKN- 161
DB 130 -KITDVISKEDEKKNPNPNDNSNSSDQKNDLQKNSDKLNDVQDEKANKENSNS 188
QY 162 NTSKSTTNPNK 174
DB 189 NDSKKNENTNK 201

RESULT 33

```

Q9UBD4 PRELIMINARY; PRT; 1419 AA.

ID Q9UBD4  
AC Q9UBD4  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE S-adenosylmethionine decarboxylase-ornithine decarboxylase.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20179918; PubMed=10713131;  
RA Mueller S., De'dera A., Luessen K., Wrenger C., Das Gupta R.,  
RA Madhubala R., Walter R.D.;  
RT "In the Human Malaria Parasite Plasmodium falciparum, Polyamines Are  
RT Synthesized by a Bifunctional Ornithine Decarboxylase, S-  
RT Adenosylmethionine Decarboxylase.";  
RL J. Biol. Chem. 275:8097-8102 (2000).  
DR EMBL; AF094833; AAF00073.1; -.  
DR HSP; P07805; 1F3T.  
DR InterPro; IPR00183; Decarboxylase2.  
DR InterPro; IPR001985; SAM decarbox.  
DR Pfam; PF02784; Orn Arg dec N; 1.  
DR Pfam; PF00278; Orn DAP Arg dec; 1.  
DR Pfam; PF01536; SAM decarbox; 1.  
DR PRINTS; PR01179; ODADCRBLASE.  
DR ProDom; PD002379; SAM decarbox; 1.  
SQ SEQUENCE 1419 AA; 166441 MW; 8842B97C05056EE6 CRC64;

Query Match 11.4%; Score 102; DB 5; Length 1419;  
Best Local Similarity 22.0%; Pred. No. 58; Matches 52; Conservative 37; Mismatches 77; Indels 70; Gaps 11;

QY 4 VQDFARNTTKEFIINKDTGEVSEL-----KPRVTYTIQNGKMSSTIVSEDP1 54  
DB 1019 VEDMSNNGFNFIYIINLGSGYPELELYDAKKGDKHHTLSLQELKDKIOKFLMBETL 1078  
QY 55 LPVY-----KGELEKGYQFDGMEISGFEGKKD-----AGYVINSKDTFIKVPKK 100  
DB 1079 KTKYGYVSPEKISLAINMSIDHY-----FSHMDNLRVICEPERYVVAASSTLAVK1134  
QY 101 -----IEKKEEENKFTF--DYSKKDNPQVNH--SQNESHRKEDLOREHSQK- 146  
DB 1135 RPTFOGIMLKEIKDHYDPLNFAQENKQDETKIHNNNDNNNNNNNNNNNNNNNOGK 1194  
QY 147 -----SDST-----KDVATVLDK--NNIS-SKSTNNPN 173  
DB 1195 GQGNIMNDLITSTNDSTSKNDHSSQVYQVNSCTIRKEDGNIKINHTINPN 1250

RESULT 34  
Q8IFNO PRELIMINARY; PRT; 1612 AA.

ID Q8IFNO  
AC Q8IFNO  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
CN PF0115C.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hamlin N., Pain A., Berriman B., Hall N., Bowman S., Churcher C.,  
RA Harris B., Harris D., Lawson D., Quail M., Bartell B.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL035477; CAD49270.1; -.  
SQ SEQUENCE 1612 AA; 190946 MW; F00A408EB4521D07 CRC64;

Query Match 11.4%; Score 102; DB 5; Length 1612;

Best Local Similarity 27.2%; Pred. No. 67;  
Matches 47; Conservative 29; Mismatches 73; Indels 24; Gaps 9;

QY 9 RNTYKFEIILNKDTGEVSELKPRVTYTIQNG-KEMSTIYSE-----EDFLPYKSEL 62  
DB 1154 RNSMNQNIILNKDTKIDFKCKENDKFKYKIKIMFPEKELNLYDGFNSNHNPM 1213  
QY 63 EKGQFDGMEISGFEGKKDAGVYINLSKDTF-IRVPFKIEKKEEENKPFDPVSKKDN 121  
DB 1214 AKESWYMW-----KKQLN-VAKYNQDYDISFKFKCIDENIKEQN---YDINQTTNN 1262

QY 122 PQVNSQLNESHKREDLOREHSQKSDSTKVATVLDKNNISSKSTNNPNK 174  
DB 1263 NTYNNQTSN--HMENLHFTDE-SSKSNNNQDKTYVT---NEVSYLPNIPKGNK 1309

RESULT 35  
Q9PPL5 PRELIMINARY; PRT; 312 AA.

ID Q9PPL5  
AC Q9PPL5  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Putative membrane protein.  
GN CJO692C.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Campylobacteriaceae; Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCITC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parthill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Raftery M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Bartell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences.";  
RL Nature 403:665-668 (2000).  
DR EMBL; AL139076; CAB72966.1; -.  
KW Complete proteome.  
SQ SEQUENCE 312 AA; 37221 MW; 0004FA7936A741B8 CRC64;

Query Match 11.3%; Score 101.5; DB 16; Length 312;  
Best Local Similarity 25.0%; Pred. No. 13; Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;

QY 15 EFLNKDTGEVSELKPRVTYTIQNGKMSSTI--VSEDPILFVYK-----GLE 63  
DB 5 DFKERQNIROMLKFSR--AINQKPLDDLRDEISSDLRRFRFKKTPNKLEIED 61  
QY 64 KGYQFDGMEISGFEGKKDAGVYINL--SKDTFIKVPKIEKKEEENKFTF--FDVSK 117  
DB 62 EYESKHTYKKSNTIYKED--LVNKLBEKOSLAKTIPSFKMERKEENKTKTKNKLFSR 118  
QY 118 KKDNP-----QVNSQLNESHKREDLOREHSQKSDSTKV--TATVLDKNNISSK--STT 169  
DB 119 KKANEIKNIQTKQTSNQAATTQTKQKKEKELNISEIKIKTEKTKIQKPLIEKKLVK 178

QY 170 NNPN 173  
DB 179 NQPN 182

RESULT 36  
Q9V7J0 PRELIMINARY; PRT; 382 AA.

ID Q9V7J0  
AC Q9V7J0  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)



RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission";  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U49945; AAC7924.1;  
 DR WormPep; C02H7.1; CE06757.  
 KW Hypothetical protein.  
 SQ SEQUENCE 535 AA; 60303 MW; 03B2D9B8BA3DFFB6 CRC64;

Query Match 11.3%; Score 101.5; DB 5; Length 535;  
 Best Local Similarity 21.0%; Pred. No. 22;  
 Matches 39; Conservative 31; Mismatches 71; Indels 45; Gaps 6;

QY 4 VQDFPANTVKEFLNK-----DTGEVSLKPHRVTVTQNGKMSSTIVSEDDILPY 58  
 DB 55 IKSAGNDKNTKATFADKLIKIDDSLNKNKAKTI---SGKDERT----- 98  
 QY 59 KGELEKGYQFDGMEISGEGKKDAGVYINLSKDTFLKPVFKIEEKEEENKPTFDVSK 118  
 DB 99 ----NKMGMCTNATSFNSRNGTG-----EKKKKKKVKEDKGDDEEEST---TKK 144  
 QY 119 KDNPNVNSQLNESHREKEDLQREHSQKSDSTK-----DVATVLDKNISSKST 168  
 DB 145 RSSKKEHHEKESKESKSAEKEKKKSSSKERHKSSDRSSEKSEKSEKSEKSEKST 204  
 QY 169 TNNPK 174  
 DB 205 TDEKPK 210

RESULT 38  
 Q9V719 PRELIMINARY; PRT; 556 AA.  
 AC 09V719;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE CG6421 protein.  
 OS ASPH OR CG6421 OR CG18658.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCB1\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe C.R., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.M., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Ffiankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Broctier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos J., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiler S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein K.C., Wu D., Yang S., Yao O.A.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Fartan D.,  
 RA Ferreira S., Frise E., Galie R.F., Gary N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jaitani M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman J., Berman B., Carlson J.W., Celniker S.B.,  
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Seale S.M.J., Smith E., Shu S., Smtunick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003808; AAF58064.2;  
 DR FlyBase; FBgn0034075; Asph.  
 SQ SEQUENCE 556 AA; 63144 MW; B420980CBDC6357A CRC64;

Query Match 11.3%; Score 101.5; DB 5; Length 556;  
 Best Local Similarity 24.5%; Pred. No. 23;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 51 EDLIPVYKGLKGYQFDG-----EISGFEKKDAGVYI-----NLKDTFLK 95  
 DB 78 EDLIPVYKGLKGYQFDG-----EISGFEKKDAGVYI-----NLKDTFLK 135  
 QY 96 PVFKIEEKEEENKPT-----PDVSKKNPNVNSQLNESHREKEDLQREHSQKSDS 149  
 DB 136 PVFKIEEKEEENKPT-----PDVSKKNPNVNSQLNESHREKEDLQREHSQKSDS 193  
 QY 150 TKDYATVLDKNISSKST 168  
 DB 194 EGTVEATVTEATTEAT 212

RESULT 39  
 Q9G082 PRELIMINARY; PRT; 765 AA.

AC 09G082;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA).  
 NC ASPH OR CG8421 OR CG18658.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20564328; PubMed=10956665;  
 RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
 RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
 RA Friedman P.A.;  
 RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved  
 RT Isoform of Asph Missing the Catalytic Domain Share Exons with  
 RT Junction";  
 RL J. Biol. Chem. 275:39543-39554(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Gawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabelarian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kesterson J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Switzkae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasatman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2195-2195(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorett V., Doup L.E., Doyle C., Drescher D., Farfan D.,  
 RA Ferriera S., Frise E., Galie R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Paclab J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Switzkae R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
 RT "Sequencing of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.B., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Seattle S.M.J., Smith E., Shu S., Smurniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
 RT "Annotation of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Flybase;  
 RT Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF289493; AAG40806.1; -  
 DR EMBL; AE03808; AAM70947.1; -  
 DR Flybase; FBgn0034075; Asph.  
 DR InterPro; IPR001440; TPR.  
 DR InterPro; IPR006025; Zn\_MTPetase.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 SO SEQUENCE 785 AA; 89843 MW; 30A8DFCD6367F1 CRC64;  
 Query Match 11.3%; Score 101.5; DB 5; Length 785;  
 Best Local Similarity 24.5%; Pred. No. 34;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 51 EPIILPVYKGELEKGVOPGW-----EISGEGKKDKGVY-----NLKDTFK 95  
 DB 78 EILDPLPLSSRSRK--VFDGWVDEHDEHGDHVEPSGEALDDHDDHDDHDEBDE 135  
 QY 96 PVFKKIEEKKKEENKPT-----PVSKKKQPNQVHNSHKKEDLQREHSOKSDS 149  
 DB 136 PLTEBELLELEEEBPTEDDEPAADBEYEDDEENNA--GENITAEDEEEBDEEDND 193  
 QY 150 TKDVTATYVLDKNNISKST 168  
 DB 194 EGTVEATVTEATTEAT 212  
 RESULT 40  
 ID Q81586 PRELIMINARY; PRT; 1377 AA.  
 AC Q81586;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 GN PFLO625C.  
 OS Plasmodium falciparum (Isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_Taxid=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlson J.W., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Chan M.-S., Nene V., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Eissen J.A., Smit A., Shalimov S.J., Sub B., Peterson J., Angiolini S.,  
 RA Perica M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M.A., Fraunholz A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,



**This Page Blank (uspto)**